DISCOVERY OF NOVEL MICROBIAL NATURAL PRODUCTS USING A CHEMICAL SCREENING APPROACH AND IDENTIFICATION OF GENE CLUSTERS RESPONSIBLE FOR BIOSYNTHESIS

A Thesis

Submitted to the Graduate Faculty

In Partial Fulfilment of the Requirements

For the Degree of

Doctor of Philosophy

Molecular and Macromolecular Sciences

Department of Chemistry

Faculty of Science

University of Prince Edward Island

Krista A. Gill

Charlottetown, Prince Edward Island

August 2016

ABSTRACT

Many of the pharmaceutical agents that are used clinically were developed from natural products and their derivatives. Natural products are often observed to have potent and specific interactions with biological targets, and are produced by a variety of macro and microorganisms. Actinomycetes have been an especially prolific source of natural products with antimicrobial applications and are responsible for approximately 70% of all clinically used agents. This has resulted in extensive investigation of these bacteria for the discovery of novel secondary metabolites, causing reisolation of known compounds to occur frequently. Novel techniques are needed to identify the production of new natural products to continue investigating natural sources for the discovery of new pharmaceutical agents.

To overcome the reisolation issue that hinders natural products discovery, a chemical based screening method was employed to prioritize strains for scale-up and further investigation. This method involved analyzing crude fermentation extracts using LC-HRMS, followed by principal component analysis to visualize the results. Utilization of this technique to analyze libraries of actinomycetes resulted in the identification of novel natural products more effectively than the tradition bioassay-guided method. This led to the selection of *Streptomyces* sp. RKKD-790 and *Kitasatospora cystarginea* NRRL B-16505 for further investigation.

A novel lasso peptide natural product tentatively named RKKD790_1592, was isolated from the fermentation broth of *Streptomyces* sp. RKKD-790. This group of peptides are characterized by their unique threaded structures, which provide them with superior stability compared to linear peptides. Computation modelling, MS/MS and stability assays were used to confirm that this novel peptide contained the characteristic threaded structure.

The strain *K. cystarginea* was found to produce three novel natural products, cystargamide, and cystargolides A and B. Unique structural features of cystargamide include the presence of the non-proteinogenic amino acids 4-hydroxyphenyl glycine and 5'-hydroxy typtophan and 2,3-epoxydecanoyl group. Cystargamide displayed the ability to activate production of pigmented metabolites by *Streptomyces coelicolor* in a disk diffusion assay. This interesting result indicated that cystargamide may be involved in bacterial communication, and may induce natural product biosynthesis in other bacteria. The gene cluster responsible for producing this novel lipopeptide was determined bioinformatically from the whole genome sequence of *K. cystarginea*. Cystargolides A and B are novel β-lactone containing peptides that were also produced by *K. cystarginea.* These peptides have very unique structures, and are the only natural products to have been described that contain a carboxyl and isopropyl substituted β-lactone group. Cystargolides A and B were found to be inhibitors of human 20S proteasome, showing that they are an excellent candidate for development as antiproliferating agents to treat cancer. Further investigations of these applications are ongoing by members of the Kerr group. The gene cluster involved in production of cystargolides A and B was predicted using bioinformatics and confirmed by generating a gene disruption mutant.

The discovery of these novel bioactive natural products represent exciting opportunities for development as an inducer of bacterial natural products, and an anticancer agent. Members of the Kerr group are continuing to explore these interesting applications. Identification of their unique structures and biosynthesis also extend our knowledge of the natural product production capabilities of bacteria.

ACKNOWLEDGEMENTS

There are several people who have contributed greatly to the completion of my studies. I would like to thank my supervisor Dr. Russell Kerr for the support, guidance and encouragement throughout my graduate studies. Thank you for accepting me as a graduate student in your lab, and for all of your guidance along the way. I would also like to thank the lab managers that have played such a prominent role in my studies. I am very thankful to have had the opportunity to learn natural products chemistry from Dr. Fabrice Berrue. Your excellent leadership, teaching skills and patience, has taught me valuable lessons in both academic and personal settings. I would also like to thank Bradley Haltli, whose passion and excitement for research is contagious. You had a lasting impact on me as a visiting undergraduate student, and I have continued to learn a great deal from you in my graduate work. I would also like to thank the other lab managers and senior members of the lab who have been very knowledgeable and helpful along the way. Dr. Dave Overy, Dr. Hebelin Correa, Dr. Gavin Carr, Dr. Malcolm McCulloch and Dr. Doug Marchbank thank you for the discussions, advice, and assistance that you have all provided.

My supervisory committee members, Dr. Christopher Kirby, Dr. Bourlaye Fofana, and Bradley Haltli have been a great support throughout my degree. Thank you for your interest in my research, your feedback that you have provided for this thesis, and throughout my entire project. I would also like to thank Dr. Pedro Quijon, who has been so helpful with every aspect of navigating through graduate school.

There are several collaborators who have contributed to the work outlined in this thesis that I would like to acknowledge. Dr. Katherine Duncan (Kerr Lab, UPEI), who isolated and characterized the RKKD library and Dr. Jennifer Arens (Kerr Lab, UPEI), who performed small scale fermentations of the *Kitasatospora* library. Patricia Boland (Kerr Lab, UPEI) and Dr. Hebelin Correa (Nautilus Biosciences Canada), who acquired LC-HRMS data, Martin Lanteigne (Kerr Lab, UPEI), who performed microwell antimicrobial assays, Nick McCarville (Nautilus Biosciences Canada) who performed antimicrobial disk diffusion assays and Kate McQuillan (Nautilus Biosciences Canada) who performed microwell cytotoxicity assays. Nadia Prigoda-Lee (Kerr Lab, UPEI) who performed the *Streptomyces coelicolor* disk diffusion assay, Bradley Haltli (UPEI and Nautilus Biosciences Canada) who performed conjugation experiments, and Maike Fischer and Lloyd Kerry (AAFC) for NMR data acquisition. Adam Proud (Pearson Lab, UPEI) and Dr. Jason Pearson (UPEI) for computational molecular modelling of KG790_1592, and Dr. Andrew Robertson and Dr. Doug Marchbank who are continuing the cystargolide medicinal chemistry project. National Cancer Institute for analysis of cystargolide B in the DTP 60-cell line assay, and Genome Quebec Innovation Centre and McGill University for genome sequencing.

I would also like to acknowledge the Nautilus staff members for all of your assistance, training, and patience. Thank you to my fellow graduate students for academic, personal and emotional support, and all of the other friends that I have met along the way!

I would also like to thank the Walker family for your encouragement and support. Thank you for being so welcoming and for treating me with such kindness. Alex, thank you for your love and support.

And finally, I would like to thank my family. Mom, Dad and Quinn, you have been my biggest supporters for everything that I undertake, and graduate school is no exception. Your enthusiasm, encouragement, and never wavering belief that I would succeed has carried me through all of the challenges that I have faced. You've been my biggest cheerleaders and despite the great distance between us, I have always felt your love and support.

Dedication

For my family

KAG, GJG, JQG

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

A	adenylation
Å	angstrom
ACN	acetonitrile
ACP	acyl-carrier protein
antiSMASH	antibiotic and secondary metabolite analysis shell
Apr	apramycin
ARS	Agriculture research services
AT	acyltransferase
BLAST	Basic Local Alignment Search Tool
bp	base pair
C	condensation
C_{18}	octadecylsilane
CDS	coding sequences
CoA	coenzyme A
COSY	correlation spectroscopy
DH	dehydratase
DMSO	dimethyl sulfoxide
DMSO-d_6	deuterated dimethyl sulfoxide
DNA	deoxyribonucleic acid
ELSD	evaporative light scattering detector
ER	enoyl reductase
ESI	electrospray ionization
EtOAc	ethyl acetate
FA	formic acid
FDAA	1-fluoro-2-4-dinitrophenyl-5-L-alanine amide
gDNA	genomic deoxyribonucleic acid
HGAP2	Hierarchical genome assembly process 2
HMBC	heteronuclear multiple bond correlation

Hpg	hydroxyphenylglycine
HPLC	high performance liquid chromatography
HRMS	high resolution mass spectrometry
HSQC	heteronuclear single quantum coherence
Htrp	Hydroxytryptophan
IC_{50}	half-maximal inhibitory concentration
ISP2	International *Streptomyces* Project medium 2
ISP3	International *Streptomyces* Project medium 3
J	coupling constant
kb	kilobase pairs
KO	knock out
KR	ketoreductase
KS	Ketosynthase
LB	lysogeny broth or Luria-Bertani
LC-HRMS	liquid chromatography high-resolution mass spectrometry
m/z	mass to charge ratio
Mb	megabase pairs
MeOH	methanol
MRSA	methicillin resistant *Staphylococcus aureus*
MS	mannitol soyflour medium
MSn	tandem mass spectrometry
NCBI	National Center for Biotechnology Information
NEB	New England Biolabs
NMR	Nuclear magnetic resonance
NOESY	Nuclear overhauser effect spectroscopy
NRP	nonribosomal peptide
NRPS	nonribosomal peptide synthetase
nt	nucleotide
ORF	open reading frame
PacBio	Pacific Biosciences

PCA	principal component analysis
PCR	polymerase chain reaction
PDA	photodiode array
PK	polyketide
PKS	polykedtide synthase
ppm	parts per million
RAST	rapid annotation using subsystems technology
RiPP	ribosomally synthesized post-translationally modified peptide
RNA	ribonucleic acid
ROESY	rotating frame overhauser effect spectroscopy
rpm	revolutions per minute
rRNA	ribosomal ribonucleic acid
R_T	retention time
SMRT	single molecule real-time
UV	ultraviolet
VRE	vancomycin resistant *Enterococcus*
WT	wild-type

CHAPTER 1: AN INTRODUCTION TO NATURAL PRODUCTS AND THEIR IMPORTANT ROLE IN DRUG DISCOVERY

1.1 Natural products and structural classifications

Natural products, or secondary metabolites are molecules that are produced by living organisms that are not required for the basic functions of life, but instead confer some kind of evolutionary advantage to the producer[1]. Natural products are produced by a variety of organisms ranging from single celled microorganisms to plants, marine organisms and mammals. As evident in Figure 1.1, the structural diversity of natural products is extremely vast, and consequently, they have been grouped together into the following classes based on structural features and biosynthetic origins: alkaloids, terpenoids, polyketides and peptides (Figure 1.1)[2,3]. Alkaloids are defined as heterocyclic nitrogen containing metabolites that are biosynthesized through extensive modification of amino acid building blocks. Alkaloids are a very complex and structurally diverse class of natural products that include subgroups like purine alkaloids (caffeine), tropane alkaloids (cocaine) and benzylisoquinoline alkaloids (morphine) to name a few. Terpenoids is the largest class of natural products to date, and contains metabolites that are derived from the 5-carbon building block isoprene. Isoprene can be produced using two distinct pathways, the mevalonate pathway, or the desoxyxyulose phosphate pathway (non-mevalonate pathway). Commonly known examples of terpenes include, paclitaxel, humulone, and β-carotene. Polyketides are metabolites that generally contain several acetate and malonate units, and differ from fatty acids in retention of oxygen containing functional groups. Polyketides are biosynthesized by large, multimodular enzymes referred to as polyketide synthases. Examples of bioactive polyketides include resveratrol, erythromycin and tetracycline. Peptide natural products contain amino acids that are linked through peptide bonds. They differ from proteins in that they are much

smaller in size, and commonly contain non-proteinogenic amino acids, modifications, and additional structural motifs including fatty acids, sugars and polyketides. Some biologically relevant examples of peptide natural products include cyclosporine, vancomycin and conotoxins. Similar to proteins, they can be biosynthesized by the ribosome to produce ribosomally synthesized post-translationally modified peptides (RiPPs) or they can be produced by large, multimodular enzymes referred to as non-ribosomal peptide synthetases (NRPSs) and be classified as non-ribosomal peptides (NRPs). Hybrid natural products that include structural features from more than one class of natural product are also observed. This creates an extreme amount of structural diversity throughout the different types of natural products.

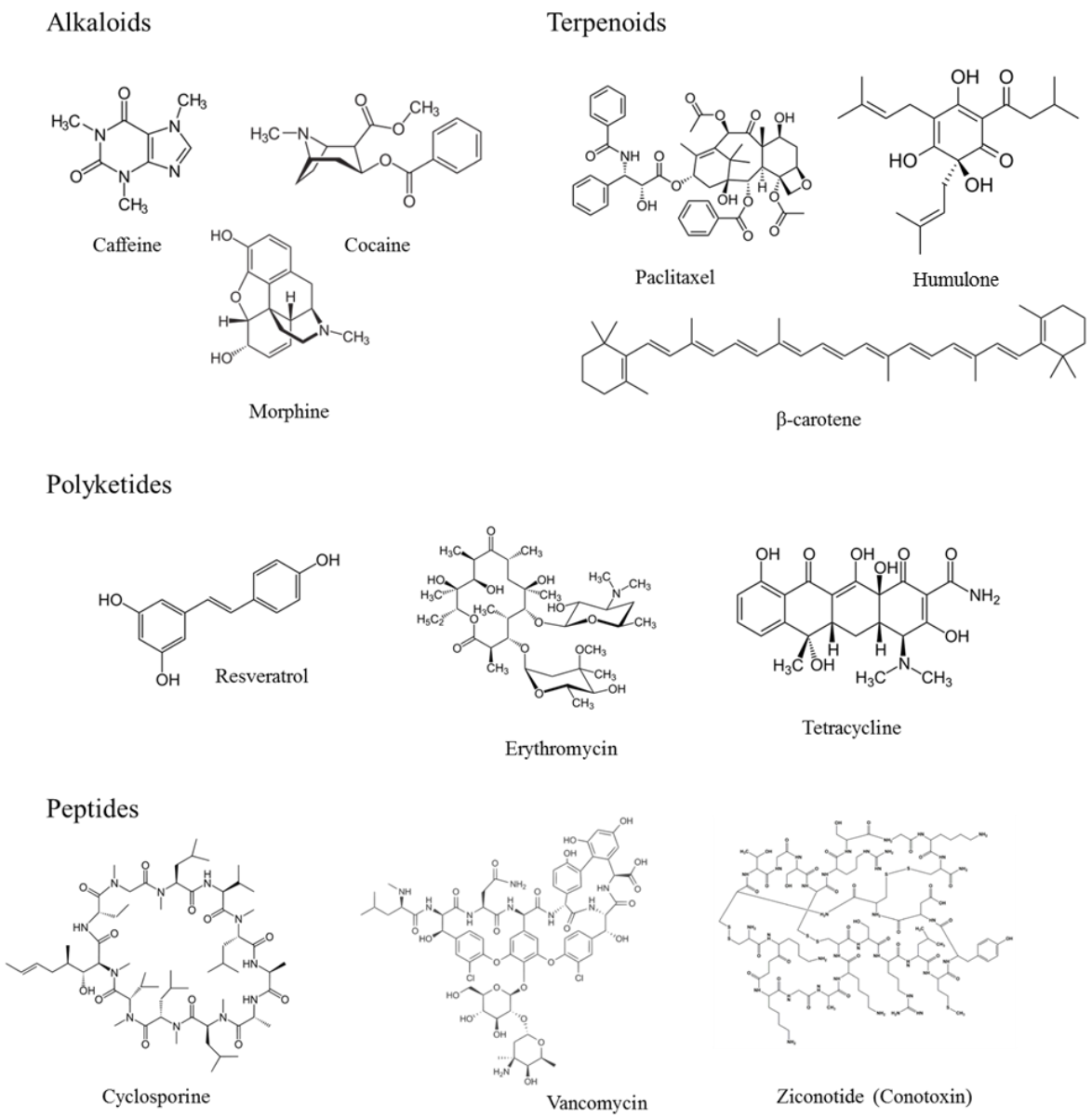


Figure 1.1. Well-known examples of natural products from each class showing structural diversity.

1.2 Ecological role of natural products

Many natural products play an important ecological role in the survival of their producing organism and provide advantages including defense mechanisms, communication, nutrient transport, and other competitive advantages[4]. Toxic metabolites are commonly produced by a variety of plant species that make the plant tissue undesirable for consumption by insects or bird predators[5]. Antimicrobial and antifungal metabolites are often produced by plants and microorganisms to protect themselves against other pathogenic microorganisms[6]. Some predatory organisms produce paralyzing agents or venoms that assist them when hunting their prey[7]. Chemical communication can be achieved through the use of natural products, and play a role in bacterial quorum sensing and regulation of gene expression[8]. Metal chelators can play an important role in nutrient sequestering and transport[9]. Lastly, natural products can act as antifouling agents and prevent microbial, plant, algae, or animal competitors from adhering to the area surrounding the producing organism[10]. The aforementioned examples represent a small sample of the many ecological roles that natural products exhibit in their producing organisms, and the true role of natural products often remains a mystery.

1.3 Biological applications of natural products

Because of the biological activity that is often displayed by natural products, they have long been exploited for their pharmaceutical applications to benefit human health[11]. Thus, natural products have played a very important role in drug discovery programs for centuries. Of the total 1 562 new drugs approved between 1981 and 2014, 4.3 % are unaltered natural products (N), 0.6 % are botanical drugs (defined mixtures) (NB), 20.5 % are natural product derivatives (ND), 3.9 % are synthetically produced with a natural

product pharmacophore (S*), 10.4 % are synthetically produced with a pharmacophore that is a natural product mimic (S*/NM), 11 % are synthetically produced with structural features that mimic the natural substrate (S/NM), 16 % are biologics that consist of large peptides or proteins (B), 6.5 % are vaccines (V), and 26.9 % are totally synthetic drugs that are not found in nature (S) (Figure 1.2)[11]. This highlights the important influence that natural products have played in the development of pharmaceutical agents.

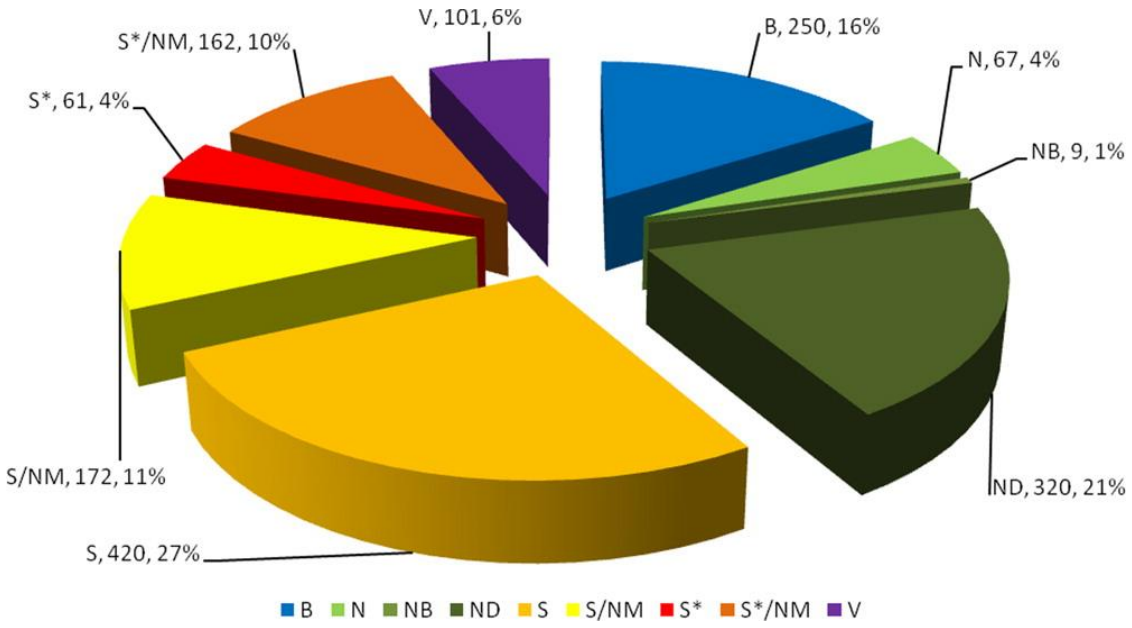


Figure 1.2. New drugs approved from 1981-2014 grouped by category of origin using the following categories:
Unmodified natural product (N), natural product derivative (ND), natural botanical (NB), biologic (B), vaccine (V), synthetic with a natural pharmacophore (S*), totally synthetic (S), natural product/substrate mimic (NM). Figure created by D.J Newman and G.M Cragg[11].

Disease treatments with the highest percentage of natural product drugs are anticancer and antibacterial agents[11]. Of the 246 anticancer drugs that were approved between 1940 and 2014, only 19 % are S, 38 % are N, ND and NB, 15 % are V and B, and the remaining 28 % are S* and /NM. If vaccine and biologics are excluded, a total of 77 % of all small molecules approved for the treatment of cancer are derived or inspired by natural products. This is an extremely significant portion of small molecule cancer treatments, and highlights the important role that natural products have played in the field of anticancer drug discovery.

Drugs that have been approved to treat bacterial infectious disease are also dominated by natural products and their derivatives[11]. Of the 141 antibacterial drugs approved between 1981 and 2014, 58 % are natural products or derivatives. Natural products also play a significant role in the treatment of other infectious diseases such as parasitic diseases. Of the 15 antiparasitic drugs, 46.6 % are natural products or derivatives. Other approved antiinfectives such as antifungal and antiviral agents include fewer natural products and their derivatives. In 2014 the World Health Organization (WHO) published a comprehensive Global Report on Surveillance of antimicrobial resistance[12]. This report highlighted the high levels of microbial resistance found in all assessed regions across the world, and included the resistance to antibiotics that are used as last resort treatment options. This shows that microbial infections may rapidly become untreatable and could result in a fatal prognosis if no new methods to combat antimicrobial resistance and infectious disease are developed. As a result, the WHO released a call to action for the development of new antimicrobial agents to combat this formidable global health concern. Since natural products have historically been the

greatest contributor to antibacterial agents, this highlights the importance of more natural products based drug development programs to help combat the growing crisis of antimicrobial resistance and untreatable infectious diseases.

Natural products are produced by organisms from all kingdoms of life. The kingdom first used by people for medicinal applications were plants, and they have continued to be fruitful sources ever since. In addition to plants, mammals, bacteria and fungi have also been producers of a significant number of natural products approved by the FDA as pharmaceuticals. Of the FDA-approved small molecule entities that are natural products or their derivatives (excluding mammalian metabolites), 45 % are derived from plant sources, 29 % from bacteria, and 22 % from fungi[13]. This shows the significant role that plant natural products have played in the treatment of a variety of human diseases including anticancer agents, analgesics and cholesterol lower agents. Taking into account only antibacterial approved drugs, the role of microorganisms becomes more significant. Of all approved antibacterial agents, 69 % are derived from natural products, and of these agents, 97 % originate from microbial sources. A total of 51 % of all antibacterial natural products are produced by bacteria and 46 % by fungi while only 3 % are produced by plants[14]. This highlights the important impact that microbial natural products have had on human health, and highlights the importance of continuous investigation of microbial sources for novel natural products.

Not only is it important to continue involving natural products discovery in drug development programs to treat cancer and infectious disease, it is also important for existing, emerging and orphan diseases. The structures of natural products have evolved over billions of years to specifically interact with molecular targets, and have an effect on

biological systems. By determining this role and biological target, natural products can continue to be exploited as pharmaceutical agents to benefit human health.

1.4 Microbial natural products

Since the discovery of penicillin in 1928 by Sir Alexander Fleming, microbes have been investigated for the production of bioactive secondary metabolites. These investigations have resulted in the identification of over 23 000 bioactive metabolites, with antiinfective, anticancer, antiviral, immunosuppressive and antiobesity activity among others[14]. Microbial natural products have had the most significant impact on clinically used antimicrobial agents, and are responsible for several classes of clinically used antimicrobial agents including the penicillins, cephalosporins, tetracyclines, aminoglycosides, chloramphenicol, macrolides, ansamycins, polyenes, and glycopeptides[15].

As stated in section 1.3, 97 % of all FDA approved antibiotic natural products and their derivatives originate from microbial sources[13]. These antibiotics are primarily produced by the the bacterial phylum actinomycetes and the fungal phyla ascomycetes and basidiomycetes[14]. *Streptomyces* is the largest genus of actinomycetes, and is the most frequent producer of bioactive natural products. *Streptomcyes* are gram-positive bacteria with relatively large genome sizes that posess the genetic ability to produce several different natural products per strain. This has made *Streptomyces* an ideal source to investigate for natural product discovery programs, which has resulted in extensive investigation of this bacterial genus[15].

1.5 Limitations of natural products discovery and strategies to overcome them

Despite the success that natural products based discovery has enjoyed for the development of anticancer and antibacterial agents, there are also several limitations. One of the major issues associated with natural products research is securing a sustainable supply. Gathering enough plant or animal tissue to purify the large quantities of natural products that are required for characterization, preclinical and clinical development and especially widespread distribution can have serious environmental impacts[16]. Synthetic and semisynthetic approaches to create a sustainable supply of natural products have seen some success, but due to the complex structures of natural products, synthetic approaches may not always be possible. Using microbes as part of natural products drug discovery programs has also been helpful for overcoming supply issues. Microbes like bacteria and fungi can be excellent producers of natural products, and unlike their plant and animal counterparts, they only need to be collected from their environment once[15]. After that they can be cultivated in laboratories for a constant and sustainable supply of the desired compound. Microbes have also been very beneficial as heterologous producers of natural products, or their precursors. Genetically engineering microbial strains to produce natural products that were isolated from other sources has been very successful at addressing issues with sustainable supplies of natural products[17].

Another issue associated with natural products research is the pace at which discoveries occur. It can be difficult to adapt natural products drug discovery in the high-throughput manner desired by pharmaceutical companies. Excellent advances have been made to generate large extract or fraction libraries, and to screen these libraries for

various biological activities, however the characterization of the metabolites in these large libraries still remains a bottleneck[18]. Increasingly sensitive analytical chemistry tools, including high resolution mass spectrometers and high-field cryoprobe NMR, have greatly assisted with improving this bottleneck, however it still remains a time consuming step in the drug development process.

A significant challenge of natural products discovery is the rediscovery of previously described metabolites. This is especially difficult when working with bacterial strains that are members of the genus *Streptomyces,* which has been extensively studied for the past several decades[19]. Many *Streptomyces* spp. share the same biosynthetic gene clusters and are able to produce the same metabolites. This causes the rate of reisolation of known natural products to be very high, especially when common bioassays like anticancer and antibacterial assays are used to screen these libraries. This creates serious issues with natural products discovery programs because it causes researchers to waste significant periods of time and resources isolating known compounds[20]. Many research groups today are focusing on the development of alternative methods to bioassay-guided isolation in order to overcome this barrier. These screening techniques involve analyses that provide more information about the chemical composition of extracts and some recently developed methods include metabolomics analyses of LC-MS[21,22] or NMR profiles[23] and molecular networking of LC-MS/MS profiles[24]. The continuation of development of these types of screening methods will help to efficiently dereplicate natural products libraries against known natural products, and help researchers prioritize hits based on chemical novelty, and eliminate or reduce the instances of reisolating known compounds.

Another way to overcome the issue of rediscovery is to investigate organisms from unique environments for the production of natural products. Organisms that have adapted to live in unique environmental niches are different than the organisms that live in commonly explored environments[25]. Exploring these unique locations can increase the likelihood of obtaining novel organisms, and therefore are more likely to produce novel natural products. There are several environmental niches that remain underexplored for natural products discovery. The marine environment is relatively underexplored compared to terrestrial environments[26]. Advances in self-contained underwater breathing apparatus (SCUBA) and especially remotely operated underwater vehicles (ROVs) have allowed researchers to access organisms and sediment samples from great depths. Cold and temperate environments are also less extensively explored for natural product discovery[25]. Tropical areas have been largely explored for natural products producing organisms because of their great biodiversity, which has left temperate and cold environments less extensively explored. Investigating organisms from such environments increases the likelihood that novel species, and therefore new natural products will be discovered.

Advances in sequencing technology and genetic engineering have revealed that many microbes have the genetic capability to produce several more natural products than what has been isolated from fermentation extracts[18]. This technology has sparked a great deal of interest in genome mining strategies for the discovery of novel natural products. These techniques can involve heterologous expression of entire gene clusters that produce putatively novel metabolites. Other techniques have involved activation of silent or cryptic gene clusters using epigenetic modifiers[27], or mimicking their natural

environment using mixed cultures[28], or environmental extracts[29]. Activation of these cryptic gene clusters will provide access to new natural products that were previously inaccessible using traditional fermentation methods. The more information that is available about biosynthetic gene clusters, and their regulation, the more successful genome based natural product discovery programs will be since it will be easier to predict the gene functions and resulting structures to find novel natural products.

1.6 Research objectives

The overall goal of this research project was to identify and characterize novel microbial natural products. In order to achieve this goal, and avoid some of the common limitations of the traditional natural products discovery pipeline, several strategies were used, which are described in Chapter 2. Firstly, to avoid downstream issues with supply and sustainability of natural products, bacterial strains were chosen as the source organisms to investigate. A library of actinobacteria were selected to investigate because historically, this group of bacteria has been the greatest microbial producers of natural products[14]. The second strategy involved selection of unique bacterial strains to avoid reisolation of known metabolites. A library of actinomycetes that were collected from marine sediment in the intertidal zone of temperate waters, was selected. A library of the less extensively studied actinomycete genus, *Kitasatospora* was also selected to increase the likelihood of identifying novel metabolites. The third strategy involved the use of a chemical based screening approach that combined liquid chromatography high-resolution mass spectrometry with multivariate analysis. This allowed for efficient dereplication of bacterial extract libraries against databases of known metabolites, and prioritization of strains producing putatively novel natural products. It also limited the number of known

metabolites that were isolated based on biological activity alone. These strategies for the discovery of novel natural products resulted in the efficient identification of three novel peptide natural products. The screening process of these bacterial libraries is described in Chapter 2, and characterization of these novel metabolites are described in Chapters 3-5.

The final aim of this research project was to sequence the genome of one of the bacterial strains that produced novel metabolites, *Kitasatospora cystarginea,* and identify the gene clusters that are responsible for the biosynthesis of these natural products. The bioinformatics analysis and gene disruption experiments used to identify these biosynthetic gene clusters are described in Chapter 6.

CHAPTER 2: LC-MS/PCA SCREENING METHOD FOR THE IDENTIFICATION OF UNIQUE METABOLOMES AMONG LIBRARIES OF ACTINOMYCETES

Some of the material included in this chapter is published as

Duncan, K., Haltli, B, Gill, K.A., Correa, H., Berrue, F. and Kerr, R.G. Exploring the diversity and metabolic potential of actinomycetes from temperate marine sediments from Newfoundland, Canada. *J. Ind. Microbiol. Biotechnol.* **2015,** *42(1),* 57-72.

Duncan, K., Haltli, B., Gill, K.A. and Kerr, R.G. Bioprospecting from marine sediments of New Brunswick, Canada: Exploring the relationship between total bacterial diversity and actinobacteria diversity. *Mar. Drugs.* **2014,** *12(2),* 899-925.

2.1 INTRODUCTION

Natural products have historically been an essential component for drug discovery programs[11]. Traditionally, the most common strategy for identifying new natural products with drug potential was bioassay-guided fractionation[30]. This process typically involved iterative rounds of chromatographic separation and bioassays until a pure, bioactive compound was identified. In the early stages of natural products research this technique proved very useful for identifying new drug candidates from natural sources. However, after several decades of natural products discovery, using this technique results in a high rate of rediscovery of known natural products. This creates a bottleneck in natural products research where identifying novel metabolites using traditional techniques becomes very challenging and time consuming.

To overcome some of the barriers of traditional natural products discovery methods, many researchers have begun using techniques that include more information about the chemical composition of extracts instead of the biological activity alone[,21,22,31,32,33]. These strategies involve untargeted metabolomics of spectrometric or spectroscopic profiles of crude extracts or fractions. Using liquid chromatography high-resolution mass spectrometry (LC-HRMS) to analyze extracts or fractions provides information about the composition of the extract provides more information about each constituent that is present including the mass, polarity, and some structural fragments. These data can then be analyzed using multivariate data analyses to visualize the results in a way that allows researchers to quickly identify putatively novel natural products.

There are several multivariate data analyses that may be employed to visualize LC-HMRS profiles of microbial fermentation extracts[31]. One effective strategy is to use

principal component analysis (PCA) for visual representation of large data sets[21,22]. Principal component analysis is a visualization technique that describes the variance among a data set by breaking down a data matrix into a set of uncorrelated variables called principal components (PCs)[34]. The first PC describes the most variance, and then each successive PC is orthogonal to the previous, and describes the next highest variance. Each data point in the matrix receives a score for each principal component, which are then plotted on a scores plot. Scores plots reveal clusters or outliers among the complex data matrix in an efficient manner. The loadings plots display the data values, in this case *m/z* ratio, that correspond to each data point in the scores plot. This technique is extremely useful for visualizing complex LC-HRMS profiles from a large library of bacterial fermentation extracts[22,21]. The use of PCA can allow for chemical dereplication of libraries by clustering of bacterial extracts that produce the same metabolites, and rapid identification of putatively novel natural products based on outliers.

LC-MS/PCA metabolomics screening is especially useful for identifying new natural products from actinomycetes[21]. Actinomycetes are an order of bacteria that have been the most prolific bacterial producers of bioactive natural products[35]. These bacteria are responsible for producing approximately one third of all clinically used antibiotics. Because of their historic importance for antibiotic discovery, this order of bacteria has been extensively mined for natural product production. This makes the issue of re-discovery extremely important, and highlights the need for new techniques to discovery novel natural products.

Streptomyces is the largest and most frequently studied genus of actinomycetes for natural product production[14,15]. They are gram positive, mycelia-forming bacteria in

the family *Streptomycetaceae*. *Streptomyces* are ubiquitous throughout the world and are commonly found growing in soil and marine sediments. Several clinically used antimicrobial agents are produced by *Streptomyces* spp. including tetracycline[36], vancomycin[37], erythromycin and more recently, daptomycin[38]. Genome sequencing of several *Streptomyces* strains has revealed the presence of 20-30 secondary metabolite biosynthetic gene clusters in each genome sequence showing that there is an even greater genetic potential to produce than what was assumed based on the number of compounds isolated from fermentations[39]. This shows that *Streptomyces* still remains an excellent source to evaluate for the discovery of new bioactive natural products, and that new techniques must be used to access these novel compounds[19].

Another closely related genus in the *Streptomycetaceae* family that is also proficient at producing natural products is *Kitasatospora*[40]. Members of this genus have very similar morphology, conserved gene sequences and natural product production abilities to their sister genus *Streptomyces*[41]. Since the first description in the 1980s, their classification has been debated, and in 1992 was altered to have all members of the *Kitasatospora* genus reclassified as *Streptomyces*[42]. The genus was re-established as a separate genus shortly after in 1997[43]. *Kitasatospora* and *Streptomyces* can be differentiated based on the stereochemistry of the amino acid diaminopimelic acid that is present in the cell wall. *Kitasatospora* contain meso and LL-diaminopimelic acid (DAP), while the cell wall of *Streptomyces* are composed primarily of only LL-DAP[40,44]. Classification of *Kitasatospora* spp. cannot be differentiated from *Streptomyces* using phylogenetic analysis of 16S ribosomal RNA sequences, and instead, multiple gene phylogenies including the β-subunit of RNA polymerase (*rpoB),* and DMP ligase (*murE)*

are required to distinguish these closely related genera[44,45]. *Kitasatospora* is comparable to *Streptomyces* in the relatively large genome sequences, and abundant presence of secondary metabolite gene clusters. Despite the characterization of only four natural products from *Kitasatospora setae*, the whole genome sequence contains 24 putative biosynthetic gene clusters[46]. *Kitasatospora* are much less abundant and less extensively studied than *Streptomyces* for the production of secondary metabolites. This shows that *Kitasatospora* spp. remain a prolific yet underexplored resource for the discovery of novel bioactive natural products.

Due to their abundant natural product production, libraries of actinomycetes were evaluated for the presence of novel compounds. Strains from the RKKD section of the Kerr Lab microbial library were isolated from marine sediment collected from Bonne Bay, Newfoundland Canada, and the Bay of Fundy, New Brunswick Canada[47,48]. The *Kitasatospora* library was obtained from the Agriculture Research Services Culture Collection, and were all originally isolated from soil samples from various locations. In this chapter, the screening of these bacterial libraries for the production of new natural products is presented. The traditional method of bioassay-guided screening was used as well as the more modern metabolomics approach of LC-MS/PCA in combination to efficiently identify novel secondary metabolite production, and prioritize strains for further investigation.

2.2 MATERIALS AND METHODS

2.2.1 Small scale fermentation of RKKD library

A total of 68 strains from the RKKD bacterial library were selected for evaluation of their natural product production by antimicrobial screening and metabolomics

analysis. All strains were streaked onto ISP2[49] agar plates and incubated for 7-10 days at 30 ºC. Single colonies were used to inoculate 10 mL of ISP2 broth containing 2-5 glass beads and incubated with agitation at 200 rpm at 30 ºC for 72 hours (seed culture). After incubation, 10 mL of fresh ISP2 was inoculated with 500 µL of seed culture and incubated under the same conditions for an additional seven days.

2.2.2 Small scale fermentation of *Kitasatospora* library

Twelve strains of *Kitasatospora* spp. were streaked onto ISP3[49] agar plates and incubated for 4-7 days at 30 ºC. A single colony from each plate was used to inoculate 10 mL of seed medium (10 g/L glucose and 10 g/L yeast extract) contained in 25 x 150 mm culture tubes containing 2-5 glass beads. Seed cultures were incubated at 30 ˚C for 48 hours with agitation at 200 rpm. A second 10 mL seed culture using the same medium was inoculated with 500 µL of stage one, and incubated under the same conditions for an additional 48 hours. A fermentation tube containing 10 mL of a lean production medium (1045 medium) consisting of 1.6 g/L dextrin, 0.8 g/L galactose, 0.8 g/L maltose, 0.8 g/L bacto-soytone, 0.4 g/L glucose and 0.3 g/L ammonium sulfate, or ISP2[49] was then inoculated with 750 µL of the second seed culture and fermented for 72 hours under the same conditions. Fermentations were repeated in triplicate for each bacterial strain in both media and un-inoculated medium blanks were included as negative controls[50,51].

2.2.3 Extraction and LC-MS analysis

Fermentation broths of all RKKD isolates and *Kitasatospora* strains as well as un-inoculated medium blanks were extracted twice with 10 mL ethyl acetate (EtOAc), and the organic fractions were combined and evaporated to dryness. Extracts were partitioned between 10 mL of 80 % aqueous acetonitrile and 10 mL of hexane. The acetonitrile phase

was removed and evaporated to dryness. The acetonitrile fractions were analyzed by LC-HRMS using a Thermo Q Exactive HPLC system with a Core Shell 100 Å C_{18} column (Kinetex, 1.7 µm 50 x 2.1 mm). The solvents used were diH_2O/0.1 % formic acid (solvent A1) and acetonitrile/0.1 % formic acid (solvent B1). A linear solvent gradient from 5 to 100 % solvent B1 over 4.8 minutes followed by an isocratic elution with 100 % solvent B1 for 3.2 minutes with a flow rate of 500 µL/min was used. Eluent was detected by ESI-MS monitoring *m*/*z* 190-2000 in positive mode, ELSD, and PDA 200-600 nm.

2.2.4 Data processing and statistical analysis

LC-HRMS files (.RAW) were converted to netCDF files using Xcalibur™ and then imported into MZmine 2[52]. MZmine was used for the following data processing steps to generate a peak list containing retention time, mass to charge ratio (*m/z)* and intensity for each metabolite detected in the LC-HRMS profiles of all of the fermentation extracts: Mass detection (1E5 threshold for RKKD library and 1E4 for *Kitasatospora* library), chromatogram building, deisotoping, alignment (mass and retention time tolerances set to 0.005 and 0.1 min respectively) and binning. Processed data was then exported as CSV files for further analysis. Artifact suppression and intensity standardization were carried out using Microsoft Excel. Intensity values above zero were replaced with a value of one, to remove bias introduced by abundant metabolites. Artifact suppression was performed by removing all features corresponding to those that were also detected in medium blanks, reserpine (reference standard), and methanol blanks from the peak lists of fermentation extracts. CSV files were then imported into The Unscrambler 10.1 (Camo Software) for data analysis. Analysis of the RKKD library and

Kitasatospora library were performed separately. Principal component analysis was used to analyze the processed peak lists using default settings provided by The Unscrambler.

2.2.5 Antimicrobial assays

All fermentation extracts from the RKKD library were tested for antimicrobial activity against the microbial pathogens, methicillin-resistant *Staphylococcus aureus* ATCC 33591 (MRSA), vancomycin-resistant *Enterococcus faecium* EF379 (VRE), *Staphylococcus warneri* ATCC 17917, and *Candida albicans* ATCC 14035. All assays were carried out according to Clinical Laboratory Standards Institute testing standards in a 96-well plate[53]. Control antibiotics used were vancomycin for MRSA and *S. warneri*, rifampicin for VRE, and nystatin for *C. albicans*. Cell growth was determined by measuring the optical density at 600 nm (OD_{600}) at time zero and after incubation for 22 hours at 37 ºC using a Thermo Scientific Varioskan Flash plate reader. The change in OD_{600} after incubation was compared to untreated negative control wells and plotted as percent growth inhibition. A single concentration of 50 µg/mL was used for each extract.

2.2.6 General fractionation and structure determination experimental procedures

Crude fractionation of fermentation extracts was achieved with reverse phase flash chromatography using a CombiFlash Rf (Teledyne ISCO) with a 15.5 g C_{18} column (High Performance GOLD, RediSep Rf). A linear gradient with diH_2O (solvent A2) and MeOH (solvent B2), that increased from 10 % to 100 % solvent B2 over 13 minutes followed by a 5 minute wash with 100 % solvent B2 and a flow rate of 30 mL/min was used.

Tandem MS analysis was performed using a Thermo LTQ Orbitrap Velos mass spectrometer. MS/MS spectra were acquired by direct infusion at a rate of 2 µL/min using an ESI source and fragmented by higher energy collisional dissociation (HCD). NMR spectra were obtained on a 600 MHz Bruker Avance III spectrometer equipped with a 1.7 mm inverse probe.

2.3 RESULTS AND DISCUSSION

2.3.1 Selection of RKKD library for screening

The library of actinomycetes that was isolated from sediment consisted of over 800 bacterial isolates. Due to the difficulty of taxonomic dereplication of actinomycete libraries, this library was classified to the genus level, and primarily consisted of closely related *Streptomyces* spp.[47,48]. The large size of this actinomycete library made fermenting and screening the entire library without high throughput equipment a time consuming and daunting task, however because of the similarity of several of the strains, this was determined to be redundant and unnecessary. Therefore, a subset of this large library consisting of 68 isolates was selected for evaluation. The subset was selected using several criteria including 16S phylogenetic analysis, presence of natural product biosynthetic gene clusters and morphology in order to optimize the possibility of identifying novel natural products. Reasons for strain selection are summarized in Table 2.1.

The main criterion for selection of the library subset was based on phylogenetic analysis of 16S rRNA sequences. A phylogenetic tree was constructed using partial 16S rRNA sequences for each isolate[47,48], and a representative isolate from each clade in this phylogenetic tree was selected to ensure a diverse representation of the library. Partial

16S rRNA sequences were compared to sequences within the NCBI database using the Basic Local Alignment Search Tool (BLASTN)[54]. Isolates that exhibited less than 98 % identity to strains in GenBank were selected. The typical threshold to delineate species of actinomyetes is a sequence identity of 99 % between 16S rRNA sequences[55], however given that partial 16S rRNA sequences were used (~500 bp), a conservative level of 98 % was used instead. Any strain whose closest BLAST relative was a rare actinomycete genus was also selected. These isolates were chosen on the basis that unique strains are more likely to produce unique metabolites. Rare actinomycetes are less extensively studied than *Streptomyces* and therefore represented a good opportunity to identify novel metabolites.

A PCR-based assay to detect the presence of secondary metabolite biosynthetic gene clusters was also used to select strains. Degenerate PCR primers that were designed to amplify conserved regions in biosynthetic gene clusters were used to screen the library for the presence of HMG-CoA reductase and dNDP-glucose 4,6 dehydratase[56]. HMG-CoA reductase is an enzyme involved in the biosynthesis of terpene natural products[57] and glucose dehydratase is involved in glycosylation of natural products[58]. Isolates that contained these biosynthetic genes and possess the ability to produce these classes of natural products were identified by the amplification of the appropriately sized PCR amplicon using these degenerate PCR primers. Strains containing these putative biosynthetic gene clusters were selected as part of the subset library.

Lastly, gross morphology and pigment production on solid media were the final criteria for selection. Pigment production was utilized as a criterion because natural products often absorb in the UV-Vis range and appear coloured, thus providing an easily

detectable indication of natural product production. Unique morphology was also used to select isolates, since unique morphology often indicates a unique strain, and may suggest differing metabolic potential and natural product production.

Table 2.1. Subset of the RKKD library that was selected for LC-MS/PCA metabolomics and antimicrobial screening, and a summary of the reasons for selection.
Several strains were selected based one or more criteria. Criteria are ordered based on most to least important from left to right respectively. NB = New Brunswick, NL = Newfoundland

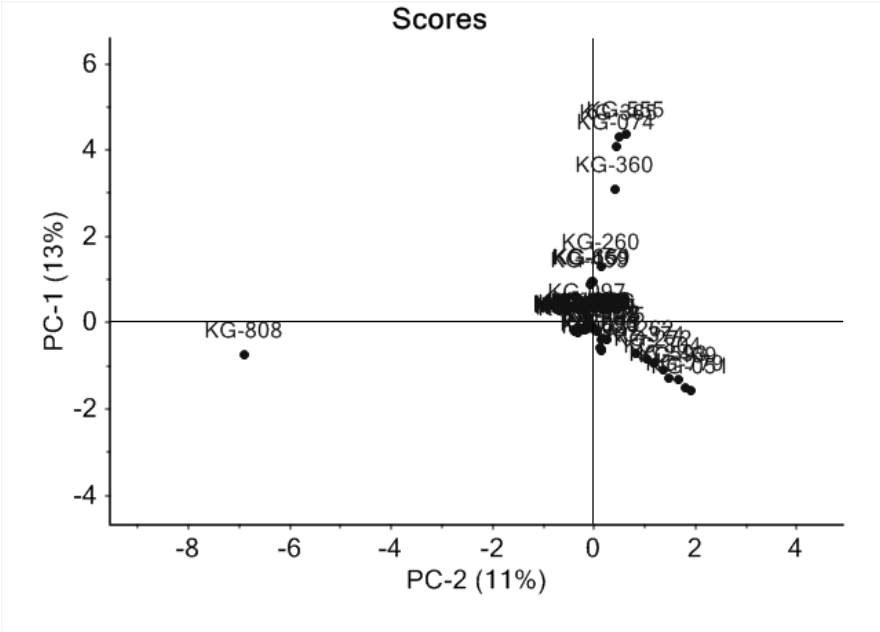
RKKD Code	Location	Selection Criteria				
		Rare Actinomycete	Sequence Similarity (<98%)	PCR Screen	Small Clade/ Clade Representative	Pigment Production/ Unique morphology
RKKD 199	NB	■		■	■	
RKKD 370	NB	■		■		
RKKD 272	NB	■			■	
RKKD 550	NL	■			■	
RKKD 686	NL	■			■	
RKKD 365	NB	■				
RKKD 012	NB		■	■		
RKKD 143	NB		■	■		
RKKD 234	NB		■	■		
RKKD 412	NB		■	■		
RKKD 097	NB		■			■
RKKD 031	NB		■			
RKKD 049	NB		■			
RKKD 075	NB		■			
RKKD 746	NL		■			
RKKD 770	NL		■			
RKKD 146	NB			■	■	
RKKD 374	NB			■	■	
RKKD 351	NB			■		■
RKKD 808	NL			■		■
RKKD 151	NB			■		
RKKD 395	NB			■		
RKKD 517	NL			■		
RKKD 563	NL			■		
RKKD 669	NL			■		
RKKD 760	NL			■		
RKKD 793	NL			■		
RKKD 799	NL				■	■
RKKD 003	NB				■	
RKKD 051	NB				■	
RKKD 059	NB				■	
RKKD 074	NB				■	

RKKD Code	Location	Selection Criteria				
		Rare Actinomycete	Sequence Similarity (<98%)	PCR Screen	Small Clade/ Clade Representative	Pigment Production/ Unique morphology
RKKD 079	NB					
RKKD 090	NB					
RKKD 123	NB					
RKKD 148	NB					
RKKD 160	NB					
RKKD 181	NB					
RKKD 194	NB					
RKKD 211	NB					
RKKD 216	NB					
RKKD 240	NB					
RKKD 260	NB					
RKKD 262	NB					
RKKD 328	NB					
RKKD 360	NB					
RKKD 403	NB					
RKKD 459	NB					
RKKD 469	NL					
RKKD 498	NL					
RKKD 516	NL					
RKKD 544	NL					
RKKD 551	NL					
RKKD 555	NL					
RKKD 593	NL					
RKKD 604	NL					
RKKD 649	NL					
RKKD 727	NL					
RKKD 736	NL					
RKKD 738	NL					
RKKD 790	NL					
RKKD 800	NL					
RKKD 804	NL					
RKKD 096	NB					
RKKD 161	NB					
RKKD 185	NB					
RKKD 667	NL					
RKKD 815	NL					

2.3.2 Prioritization of the RKKD library using multivariate data analysis and antimicrobial activity.

All 68 isolates that were selected for metabolomics analysis were fermented, extracted and analyzed by LC-HRMS. Chemical profiles were processed and analyzed by PCA to efficiently identify strains that produced unique metabolites. Due to the large size of the resulting data matrix, a high threshold of 1E5 was used during peak detection in order to ensure that only abundant metabolites contributed to prioritization of strains for isolation of metabolites. A total of 366 distinct mass features were detected using MZmine 2. Of those features, 16 were detected in MeOH blanks, media blanks, or internal reserpine blanks (calibration standard) and were removed from the features list. Peak intensities were normalized to a value of one for all detected features to remove bias associated with abundance. Principle component analysis was then used to visualize the processed data matrix using The Unscrambler. Scores plots of PC1 vs. PC2 to PC15 vs. PC16, shown in Figure 2.1, were used to analyze the metabolomes of isolates from the RKKD library. Based on the scores plots, 10 strains clustered away from the other strains, suggesting the presence of mass features that are not observed in the majority of strains. The following initial hits were identified: RKKD-808 from PC1 vs. PC2; RKKD-240 from PC3 vs. PC4; RKKD-161 from PC5 vs. PC6; RKKD-412 and RKKD-760 from PC9 vs. PC10; RKKD-790, RKKD-555 and RKKD-760 again from PC11 vs. PC12; RKKD-074, RKKD-374 and RKKD-790 again from PC13 vs. PC14; RKKD-234 from PC15 vs. PC16.

A



B

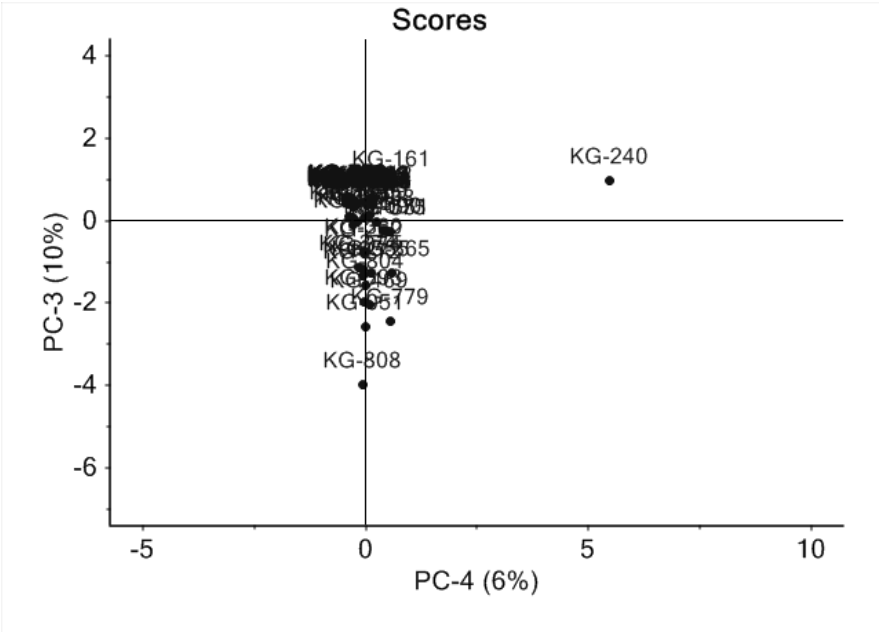
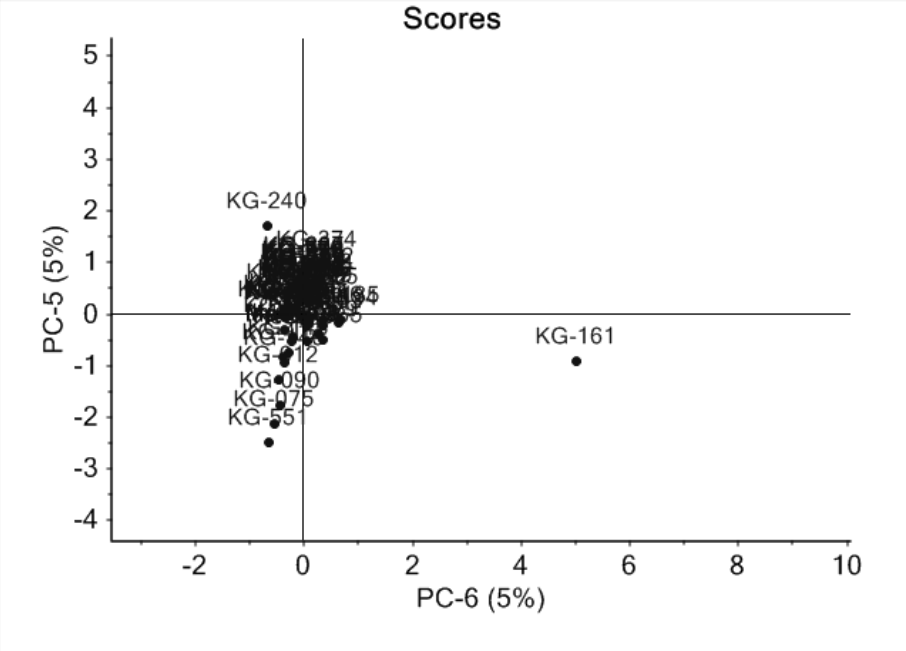


Figure 2.1. Principal component analysis scores plots of RKKD library in ISP2 medium.

C



D

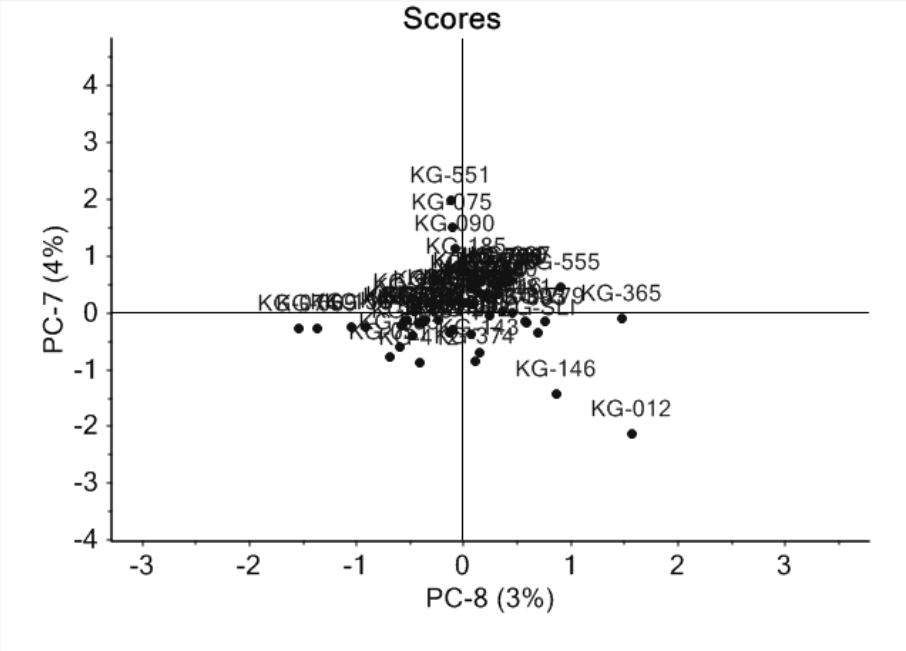
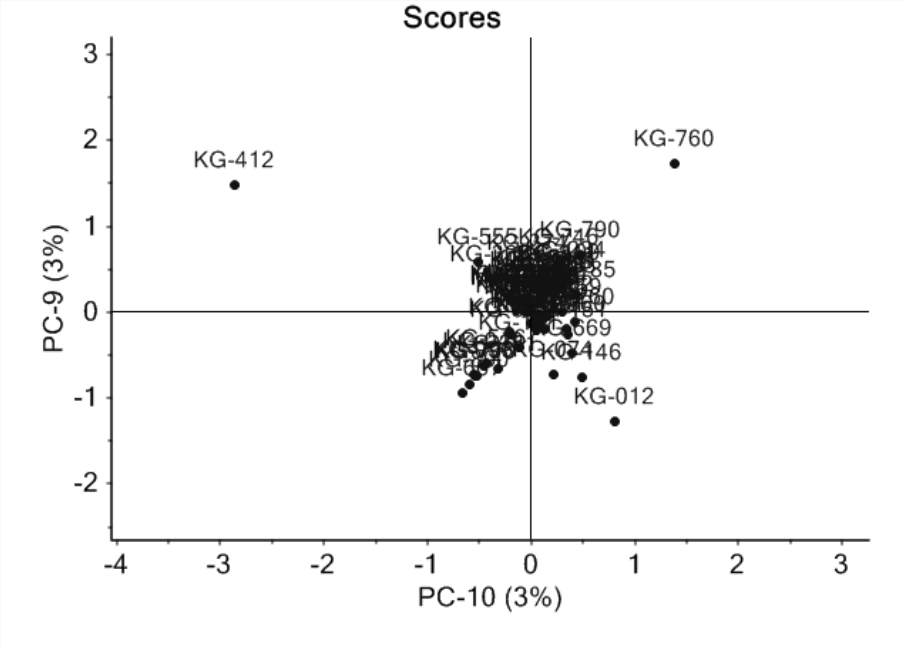


Figure 2.1. Continued from previous page.

E



F

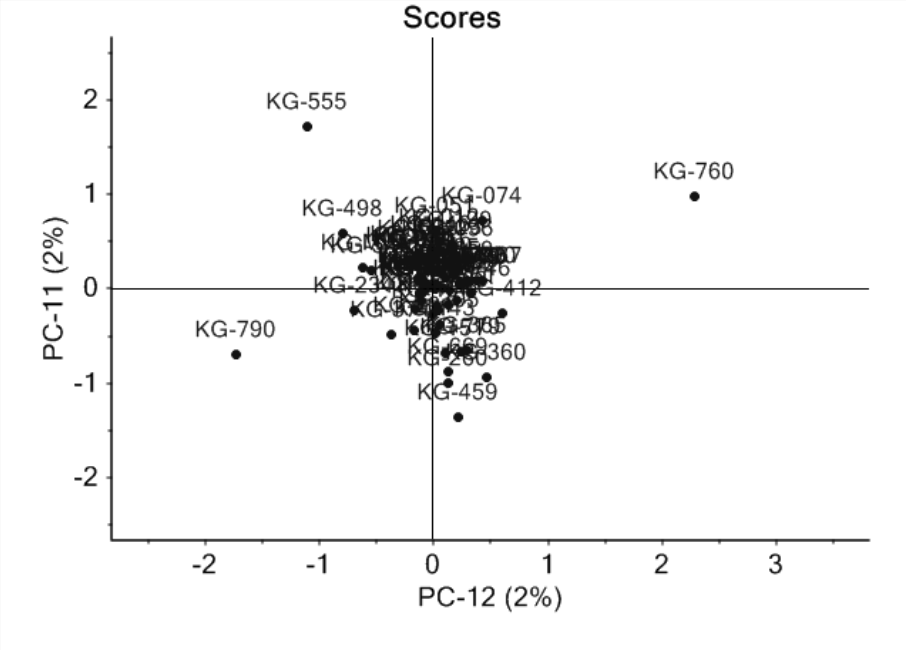
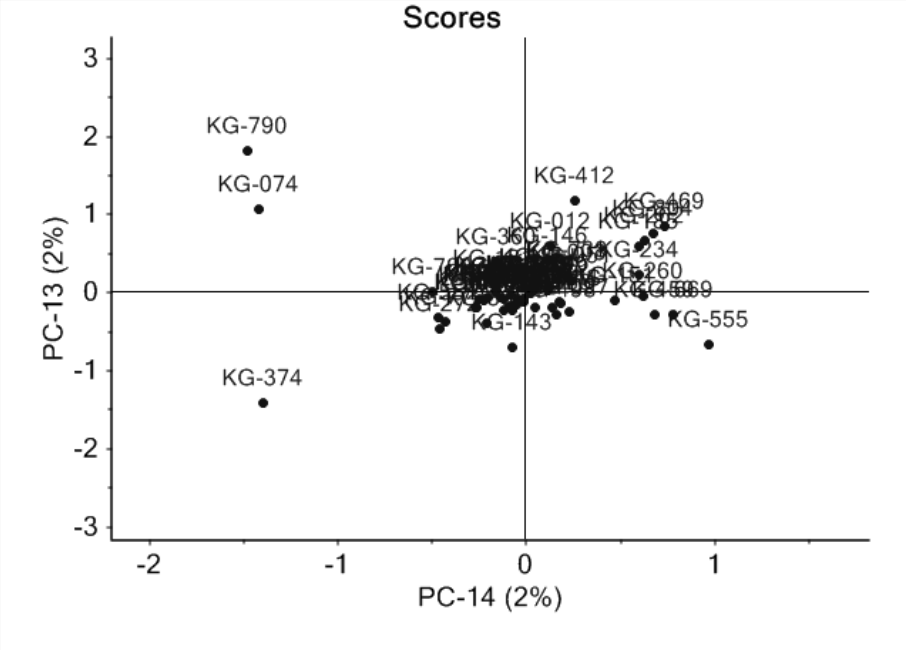


Figure 2.1. Continued from previous page.

G



H

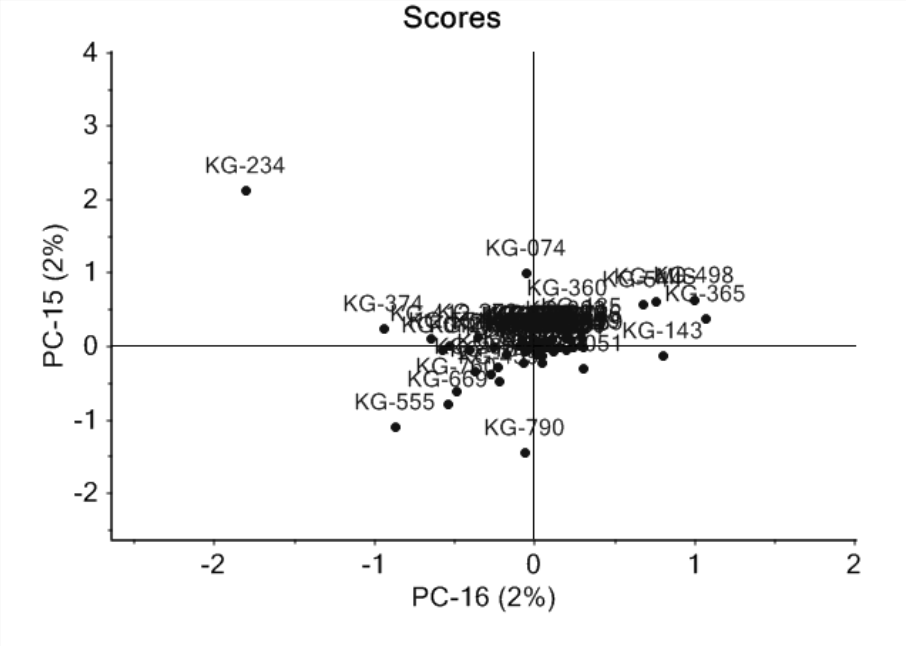


Figure 2.1. Continued from previous page.

The fermentation extracts of all 68 isolates from the RKKD library were investigated for antimicrobial activity against four bacterial pathogens and one fungal pathogen. The results of these bioassays are shown in Figure 2.2, and strains displaying activity are summarized in Figures 2.3 and 2.4. Extracts were considered active if they inhibited the growth of one or more of the microbial pathogens by greater than 70 %. Based on the results from the antibacterial assays, 11 strains were identified as producers of metabolites with antibacterial activity. Eight strains inhibited the growth of both MRSA and VRE (RKKD-059, RKKD-074, RKKD-123, RKKD-240, RKKD-260, RKKD-412 and RKKD-808), three strains inhibited the growth of only MRSA (RKKD-090, RKKD-234, and RKKD-498), and one strain inhibited the growth of only VRE (RKKD-790). It is worth noting that RKKD-234 showed moderate growth inhibition against VRE as well, but lower than the threshold selected to be classified as a hit. RKKD-790 also displayed weak growth inhibition against MRSA, and RKKD-412 also inhibited the growth of *P. vulgaris* but at levels below the threshold.

The antifungal activity against *C. albicans* was also evaluated. Antifungal hits were selected based on growth inhibition of 70 % or greater against *C. albicans*. A total of 12 antifungal hits were identified as follows: RKKD-051, RKKD-148, RKKD-234, RKKD-262, RKKD-272, RKKD-374, RKKD-412, RKKD-469, RKKD-593, RKKD-779, RKKD-804, and RKKD-808 all inhibited the growth of *C. albicans*.

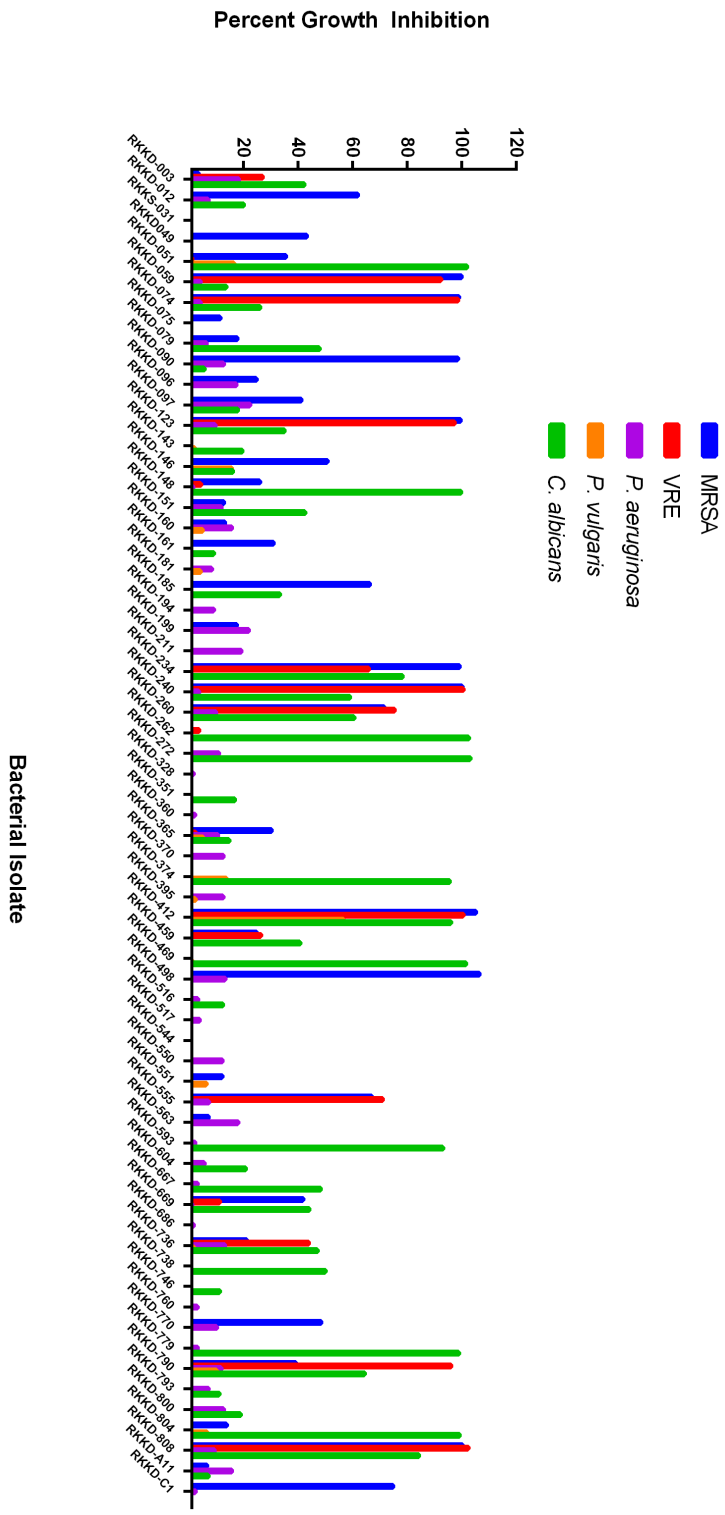


Figure 2.2. Percent growth inhibition of RKKD subset library against five microbial pathogens

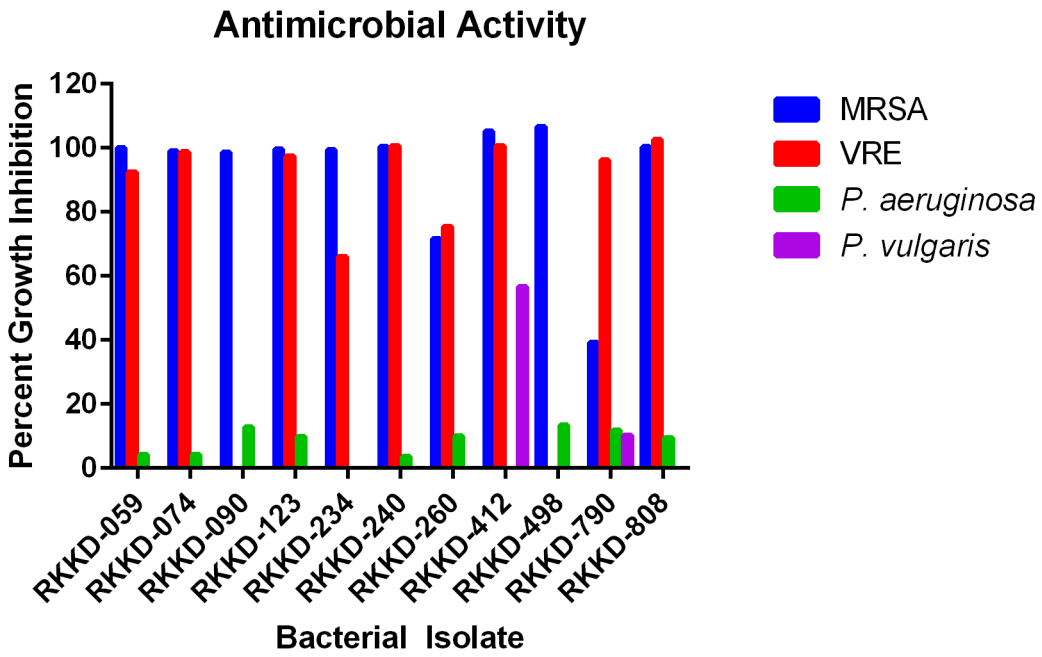


Figure 2.3. Summary of the percent growth inhibition of extracts from the RKKD library that displayed activity against four human bacterial pathogens.

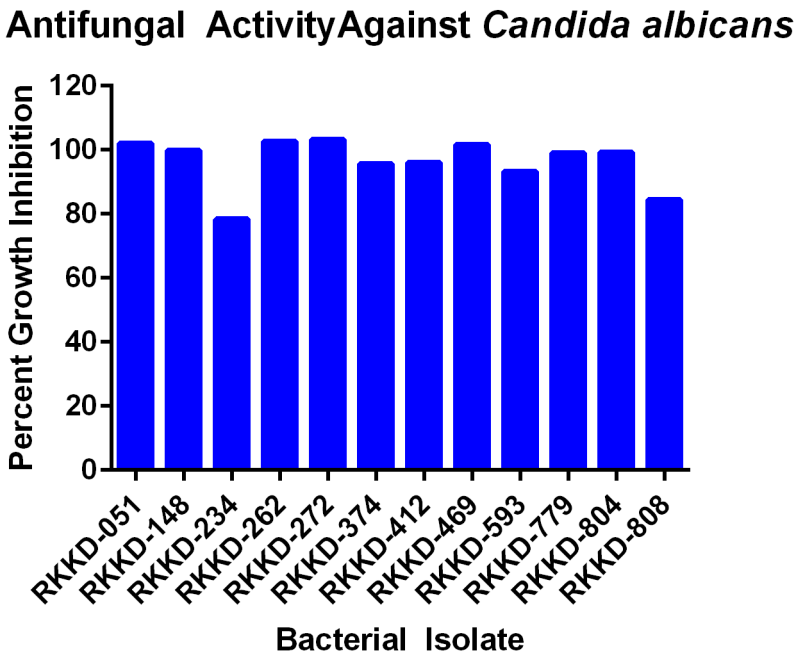


Figure 2.4. Summary of the percent growth inhibition of extracts from the RKKD library that displayed activity against *Candida albicans*.

Some strains were identified as hits by more than one of the bioassays or metabolomics screen. A total of three strains were identified as hits using all three screening methods: RKKD-234, RKKD-412 and RKKD-808. These strains were given the highest priority. Four strains were selected as hits in the LC-MS/PCA screen, and also displayed either antibacterial or antifungal activity and were classified as the next highest priority. Three of the strains were considered hits from the metabolomics screen but did not have any antimicrobial activity. Lastly, there were 13 strains that were active in either the antibacterial or antifungal assay alone. Strains that were hits in a single screening method were given the lowest priority.

To further prioritize the strain list, raw LC-MS data was used to assess the putative novelty and relative abundance of the metabolites each strain produced using the ELSD chromatogram to determine the practicality of scale-up fermentations and compound purification. After analyzing the LC-MS chromatograms of each prioritized strain, some were removed from the list based on observations from the ELSD chromatogram. The ELSD signal can show relative abundance of each metabolite. Some strains were deprioritized because they produced a very low abundance of the metabolites detected by MS, and it would be very challenging to obtain enough material for full characterization. The strains that were deprioritized because of low abundance included: RKKD-059, RKKD-123, RKKD-148, RKKD-262, RKKD-498, RKKD-593 and RKKD-804.

Several strains were also grouped together because they produced the same metabolites. RKKD-051, RKKD-090, RKKD-272, RKKD-374, RKKD-469 and RKKD-779 all produced the same two metabolites with exact masses of *m/z* 912.6269 $[M+H]^+$

(RKKD_912) and *m/z* 445.1140 $[M+H]^+$ (RKKD_445). At the time, Hebelin Correa another member of the Kerr lab was investigating other strains from the RKKD library for the production of novel metabolites and found several strains that also produced these same metabolites. She was working on determining the identity of these compounds, therefore, all of these strains were removed from this priority list. Upon full characterization she determined that RKKD_445 was enterocin[59] and RKKD_912 was the recently described metabolite, reginamide A[60].

Three other strains were grouped together because of similar metabolite production. RKKD-074, RKKD, 555 and RKKD-260 all produced two metabolites with *m/z* 828.4741 $[M+NH_4]^+$ and two metabolites with *m/z* 870.4822 $[M+NH_4]^+$. Upon searching these masses in AntiBase, RKKD_828 corresponded to eight known compounds. The rationale for deprioritizing these strains was that they had a lower likelihood of producing novel compounds because their metabolite profiles were not unique among this relatively small strain library even though they were isolated from different locations. There were also several known compounds in the database that corresponded to the exact mass. The aim of this investigation was to identify metabolites with novel and unique structures, therefore these strains were deprioritized.

To further prioritize strains by dereplicating against known natural products, the pseudomolecular ion for each metabolite was determined, and database searches using AntiBase 2007 and SciFinder Scholar were used to determine if the exact mass corresponded to any previously described natural products. Because of the accuracy of the high resolution MS data, metabolites could be identified as known compounds or putatively novel compounds with relatively high confidence. Many strains were

deprioritized based on the putative production of known natural products. Additional information like the producing organism, UV absorbance, MS/MS fragmentation, and NMR were also used to confirm the identity of these known compounds.

Table 2.2. Summary of hits from the LC-MS/PCA screen and antimicrobial screens showing the criteria for prioritizing and deprioritizing strains.

Strain	Higher Priority				Lower Priority		
	Antibase 2007	PCA	Antibacterial Activity	Antifungal Activity	Low Abundance	Collaborators	Similar Profiles
RKKD-790	■	■	■				
RKKD-240	■	■	■				
RKKD-161	■	■					
RKKD-412		■	■	■			
RKKD-808		■	■	■			
RKKD-234		■	■	■			
RKKD-760		■					
RKKD-123			■		■		
RKKD-498			■		■		
RKKD-059			■		■		
RKKD-148				■	■		
RKKD-262				■	■		
RKKD-593				■	■		
RKKD-804				■	■		
RKKD-374		■		■		■	
RKKD-090			■			■	
RKKD-051				■		■	
RKKD-272				■		■	
RKKD-469				■		■	
RKKD-779				■		■	
RKKD-074		■	■				■
RKKD-555		■					■
RKKD-260			■				■

2.3.3 Identification of metabolites from RKKD library

RKKD-234 was selected because of its observed antimicrobial activity in both the antibacterial and antifungal assays, and the LC-MS/PCA chemical screening method. The LC-MS chromatogram showed that RKKD-234 produced a single metabolite in high abundance, and was labelled RKKD234_467 (Figure 2.5). The pseudomolecular ion was determined to be *m/z* 467.2088 $[M+H]^+$ and upon searching in AntiBase, this query returned the natural product staurosporine with a mass difference of 2.14 ppm compared to the predicted mass (*m/z* 467.2078 [M+H]+ calc for $C_{28}H_{27}N_4O_3^+$). This value is within the accepted 5 ppm variance for HRMS. NMR spectra acquired with the crude extract were used to confirm that the structure of RKKD234_467 was the same as staurosporine (Appendix Figure 8.1). Staurosporine was the first member of the bis-indole family of alkaloid natural products to be described in the literature[61,62]. It was originally isolated from *Streptomyces staurosporius,* although it has been isolated from several different *Streptomyces* strains since its initial isolation[62]. It has a diverse array of reported biological activities, however the most extensively studied is inhibition of protein kinase, which results in broad cytotoxic activity[63]. The observed antibiotic and antifungal activity of RKKD234_467 corresponded to the bioactivity reported for staurosporine and further supported the identification of RKKD234_467 as staurosporine.

RKKD-412 was also a strain that was identified as a hit using all three screening methods. The LC-MS data revealed the production of one major and one minor metabolite (Figure 2.6). Fractionation of the crude extract led to identification of a metabolite RKKD412_613 with *m/z* 613.2413 $[M+H]^+$. Database queries of this exact mass showed that the structure of RKKD412_613 likely corresponded to the known

natural product, novobiocin. Novobiocin is one of the most commonly isolated natural products from many *Streptomyces*, and is an aminocoumarin natural product that has potent antimicrobial activity[64,65]. This metabolite was likely responsible for the antimicrobial activity that was observed for the crude extract from RKK-412. Novobiocin was first isolated from the strain *Streptomyces niveus* and has been referred to by several other names including streptonivicin, albamycin and cathomycin[66].

The third and final strain that had positive results in all three screening methods was RKKD-808. The most abundant metabolite produced had an exact mass of *m/z* 765.4832 $[M+H]^+$ (Figure 2.7). This strain produced several closely related metabolites that all differed by 14 mass units, which likely corresponded to the presence or absence of additional methylene groups. These metabolites were identified from their exact masses as members of the nactin family of cyclic polyether natural products[67]. This family of natural products is commonly produced by several *Streptomyces* strains and have been reported as antimicrobial and cytotoxic agents[68]. Due to the similarity in exact masses of several of these metabolites, similar observed bioactivity, and isolation source, the metabolites produced by RKKD-808 were tentatively determined to be members of the nactin family of natural products. This strain was not investigated further for its metabolite production.

The strain RKKD-760 was prioritized based on results from the chemical profile screen. This strain produced two major metabolites with exact masses of *m/z* 825.4401 $[M+H]^+$ and 839.4579 $[M+H]^+$ (Figure 2.8). Database searches of the pseudomolecular ion 825 matched the exact mass of arylomycin A2 and A3 within 0.9 ppm[69]. The aryolomycins are lipopeptide natural products that were isolated from *Streptomyces* sp.

Tu 6075. Due to the similarity in molecular weight and producing strains, RKD760_825 was tentatively assigned as arylomycin. RKKD760_839 had a mass difference of 14 compared to RKKD760_825, which corresponds to the mass of an additional methylene. Because of the similarity in retention time, UV spectrum and mass difference corresponding to an additional methylene compared to RKKD760_825, RKKD760_839 was tentatively assigned as an arylomycin analog. This strain was not investigated any further for its natural product production.

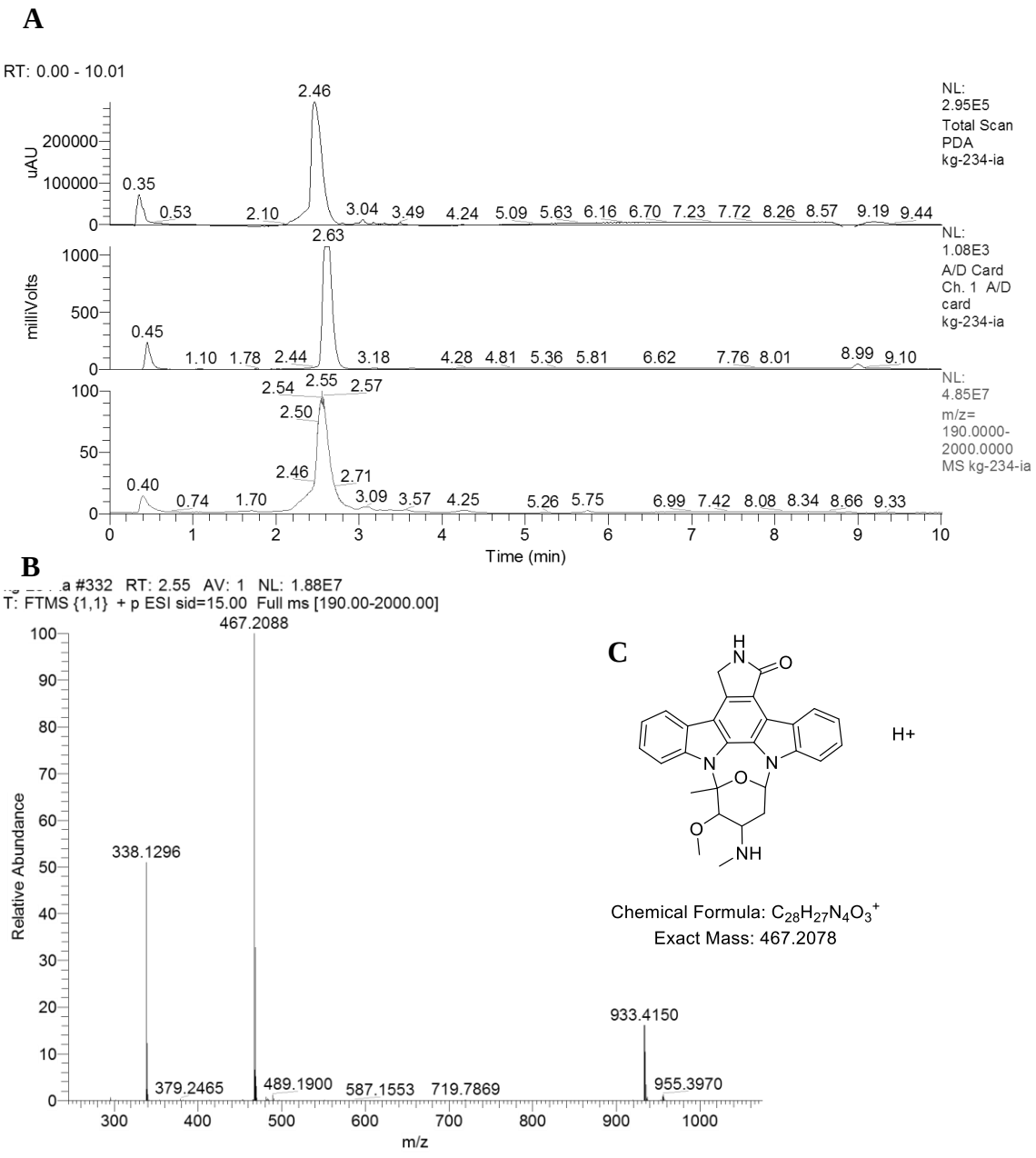
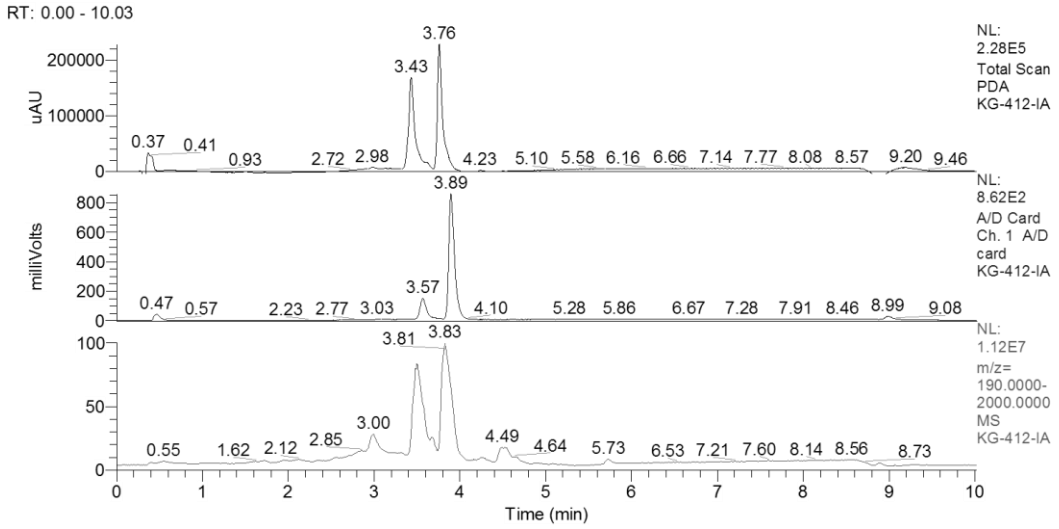
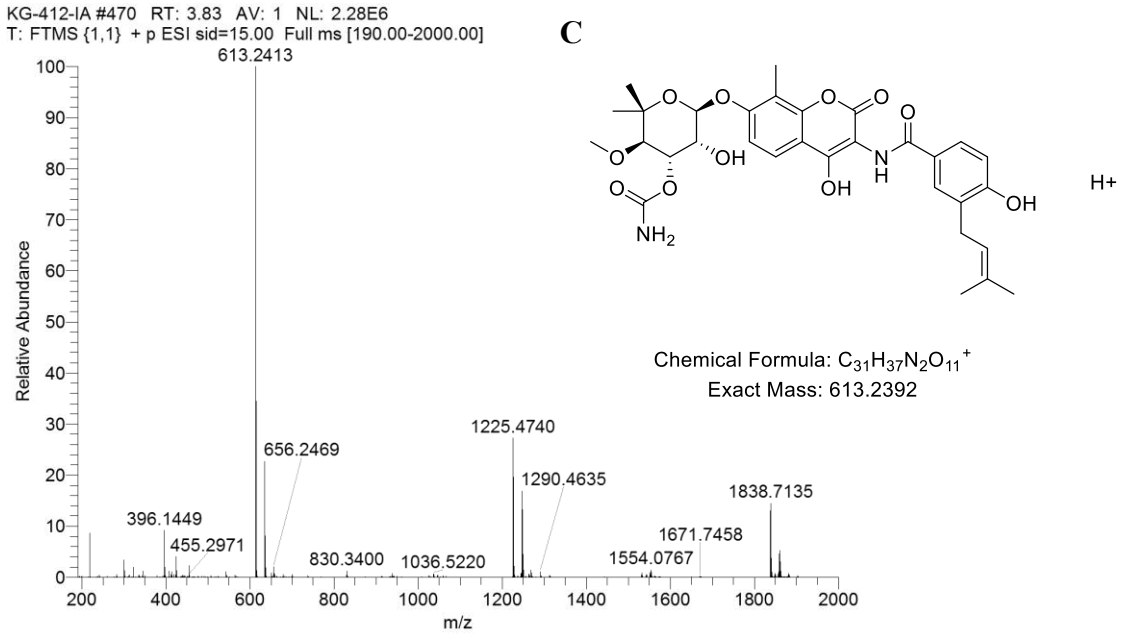


Figure 2.5. **A**. Chromatograms of the crude extract from RKKD-234 grown in ISP2. Three detectors were used to generate 3 chromatograms for the same sample: PDA, ELSD and +ESI-MS. **B**. The high resolution mass spectrum of RKKD234_467. **C**. The molecular structure of staurosporine.

A



B

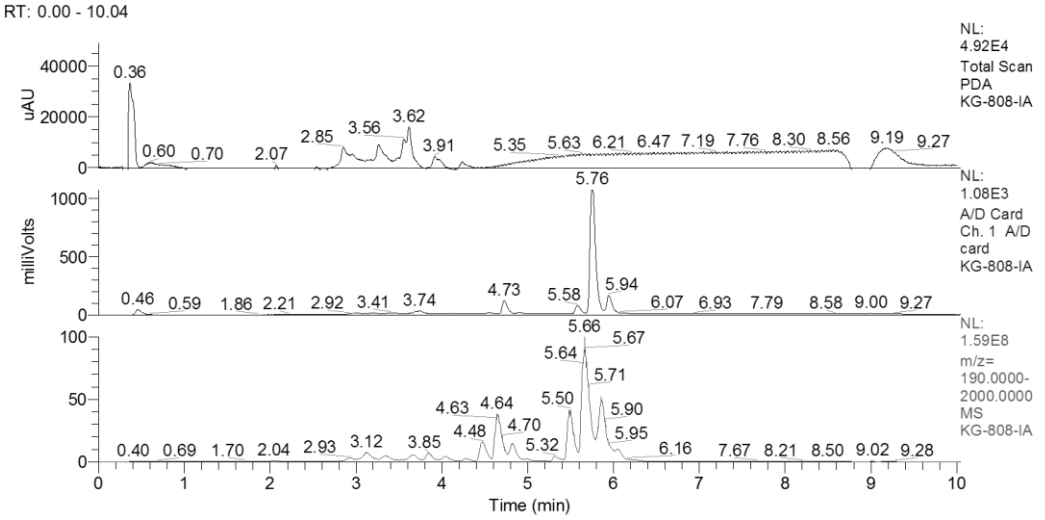


C

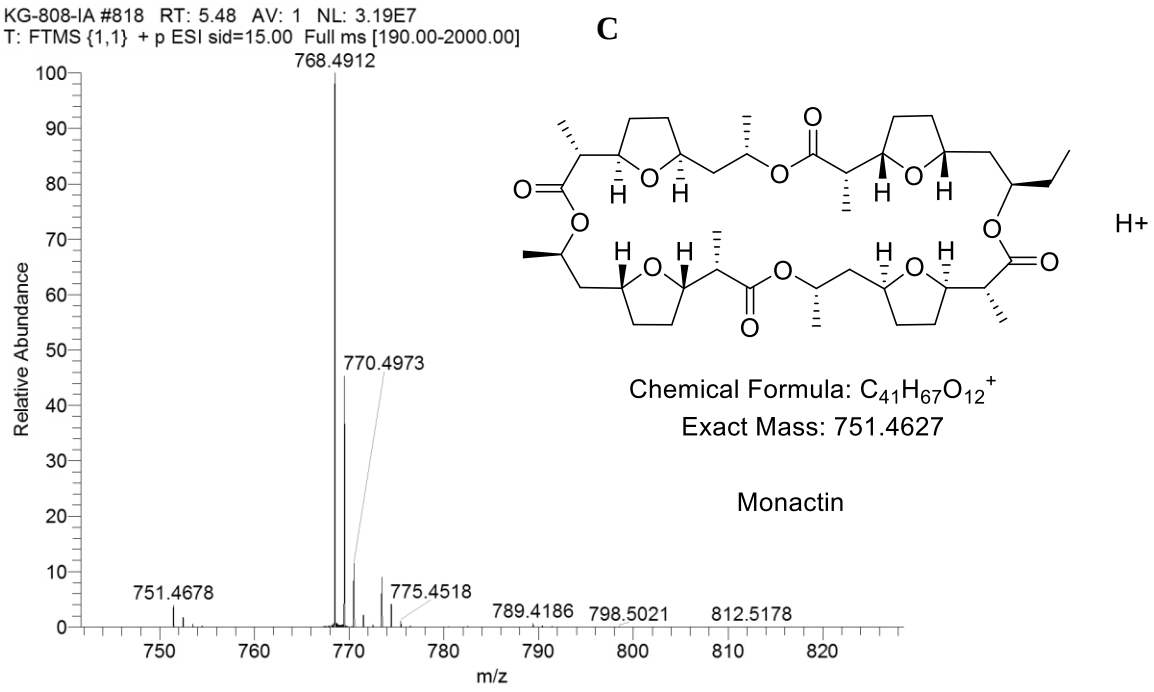
Chemical Formula: $C_{31}H_{37}N_2O_{11}^+$
Exact Mass: 613.2392

Figure 2.6. **A**. Chromatograms of the crude extracts from RKKD-412 grown in ISP2. Three detectors were used to generate 3 chromatograms for the same sample: PDA, ELSD and +ESI-MS. **B**. The high resolution mass spectrum of RKKD412_613. **C**. The molecular structure of novobiocin.

A



B



C

H+

Chemical Formula: $C_{41}H_{67}O_{12}^{+}$

Exact Mass: 751.4627

Monactin

Figure 2.7. **A**. Chromatograms of the crude extracts from RKKD-808 grown in ISP2. Three detectors were used to generate 3 chromatograms for the same sample: PDA, ELSD and +ESI-MS. **B**. The high resolution mass spectrum of RKKD808_751 **C**. The molecular structure of monactin **D**. The high resolution mass spectrum of RKKD808_765. **E**. The molecular structure of dinactin.

D

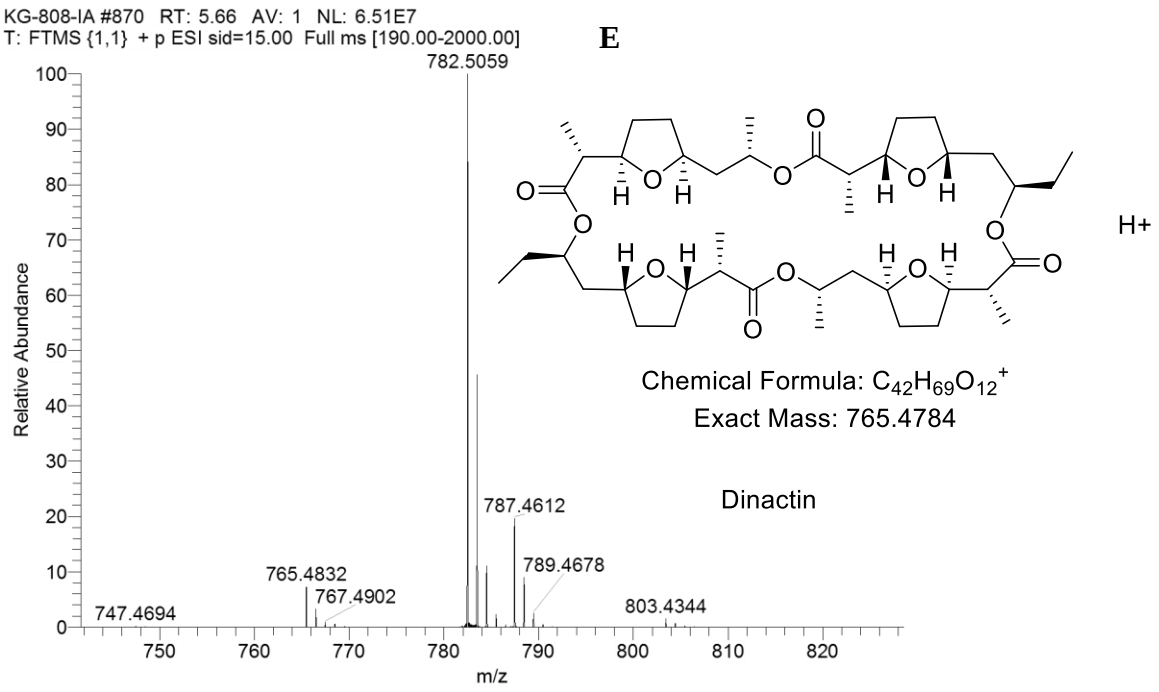


Figure 2.7. Continued from previous page.

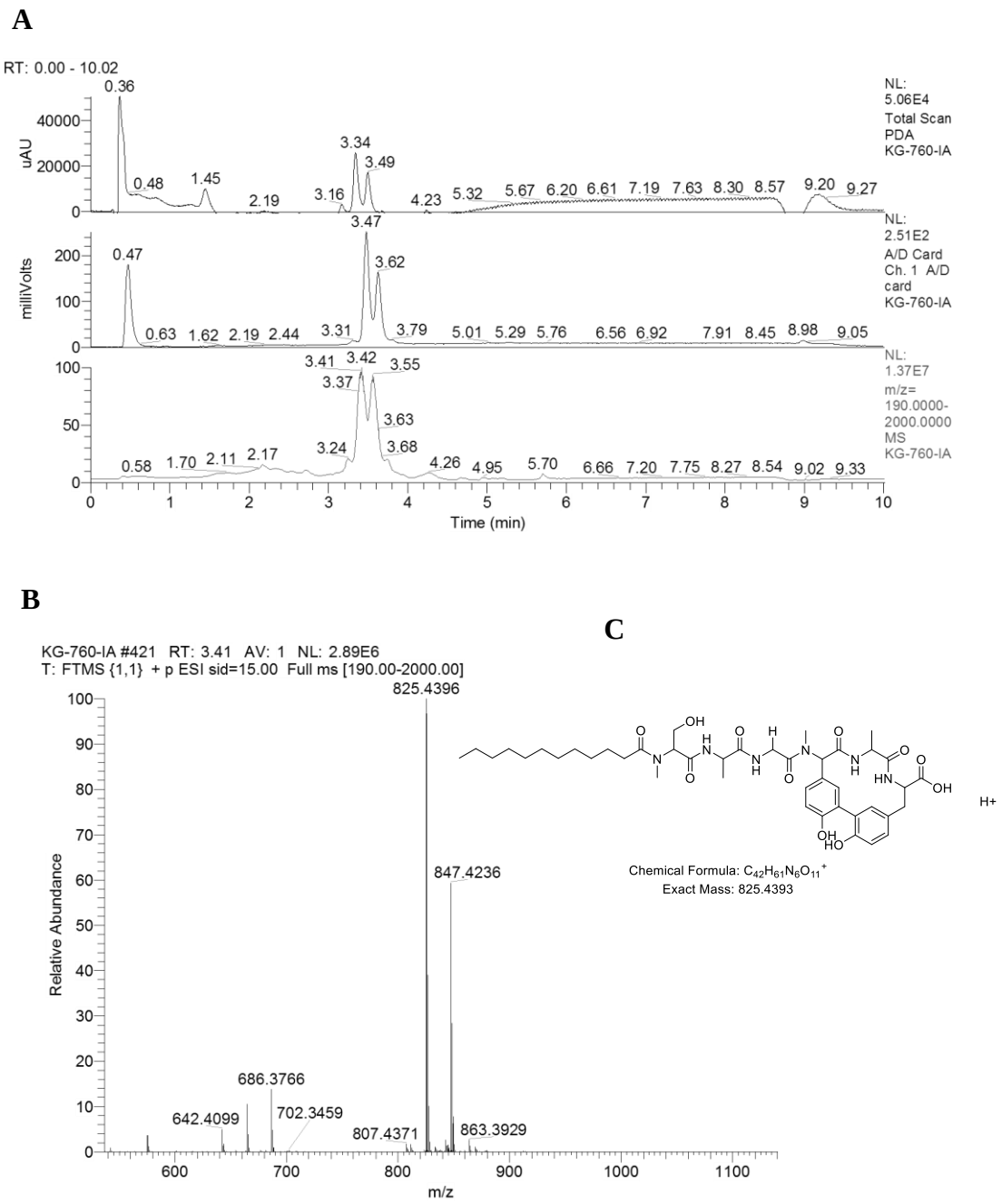


Figure 2.8. **A**. Chromatograms of the crude extracts from RKKD-760 grown in ISP2. Three detectors were used to generate 3 chromatograms for the same sample: PDA, ELSD and +ESI-MS. **B**. The high resolution mass spectrum of RKKD760_825 **C**. The molecular structure of the arylomycin A2.

After deprioritizing the strains RKKD-234, RKKD-412, RKKD-808 and RKKD-760 based on depreplication against known natural product databases, three strains remained for investigation. The strain RKKD-161 was initially prioritized because of its unique chemical profile. When a large scale fermentation extract of this strain was analyzed by LC-MS, the metabolite profile was different than what was observed for the initial small scale extract. Since the fermentation results were not reproducible, this strain was deprioritized. Efforts were focused on investigating other strains that produced putatively novel compounds and therefore this strain was not investigated any further for its natural product production. This example highlights the importance using replicate fermentations when analyzing the metabolite production of a library of bacterial strains. All subsequent small-scale fermentations that were used for LC-MS/PCA screening for the identification of novel metabolites were performed in triplicate.

RKKD-240 was a strain that was prioritized because of its unique chemical profile and antibacterial activity against MRSA and VRE. This strain was fermented on a larger scale, and fractionated to identify the metabolites that were produced. This resulted in the isolation of the metabolite labelled RKKD240_401 which had an exact mass of *m/z* 401.2678 $[M+H]^+$. The MS fragmentation pattern of RKKD240_401 showed similarity to a minor metabolite present in a different fraction that was labelled RKKD240_563, which had an exact mass of *m/z* 563.3216 $[M+H]^+$ (Figure 2.9). At the time of isolation, no known natural products in Antibase (2007) corresponded to this exact mass. Both MS spectra contained the fragment *m/z* 383.2580, which would correspond to the fragments $[M-H_2O+H]^+$ and $[M-C_6H_{12}O_2+H]^+$ for RKKD240_401 and RKKD240_563 respectively. This suggested that RKKD240_563 and RKKD240_401 had the same parent structure

differing in that RKKD240_563 was glycosylated and RKKD240_401 was hydroxylated. The NMR spectra of RKKD240_401 revealed that the structure was the same as the recently described actinopyranone natural product PM050463[70] (Appendix, Figure 8.2). The glycosylated analog PM050511 had the same parent structure with the addition of D-glucose, and matched the exact mass of RKKD240_563. These metabolites were isolated after 2007, and consequently were not present in the 2007 version of AntiBase. These natural products were isolated from *Streptomyces albus* POR-04-15-053, which like *Streptomyces* sp. RKKD-240 was also cultured from marine sediment. These metabolites were reported to have cytotoxic activity against breast (MDA-MB-231), colon (HT29) and non-small cell lung (A549) cancer cell lines. Other members of the actinopyranone family of natural products have been isolated from different strains of *Streptomyces* and are reported as antimicrobial and cytotoxic agents[71,72].

RKKD-790 was selected for further investigation of its metabolite production due to observed antibacterial activity against VRE, and its unique metabolome identified by LC-MS/PCA. RKKD-790 produced two distinct metabolites: RKKD790_1128 and RKKD790_1592 with exact masses of *m/z* 1128.6648 and 1592.8184 (Figure 2.10). Database searches revealed no previously identified natural products for RKKD790_1592. Fractionation of a large-scale extract from RKKD-790 revealed that RKKD790_1128 was responsible for the observed antibacterial activity. Closer investigation of the mass spectrum of purified RKKD790_1128 revealed that the base peak with an exact mass *m/z* 1128.6648 was the ammonium adduct and *m/z* 1111.6466 was the proton adduct. The structure of RKKD790_1128 was determined to be valinomycin using tandem MS (Appendix, Figure 8.3) [73,74]. Valinomycin is a cyclic

peptide that has reported antimicrobial activity, similar to that observed for RKKD790_1128 against VRE[75]. The putatively novel compound, RKKD790_1592 was also purified from the large-scale extracts of RKKD-790. Although fractions containing this metabolite were not active in the antibacterial assay, it was purified because of its putative structural novelty. The structure of RKKD790_1592 was solved using extensive 1D and 2D analysis and was determined to be a lasso peptide that was a novel metabolite at the time of isolation. The full characterization of this new natural product is discussed in Chapter 3.

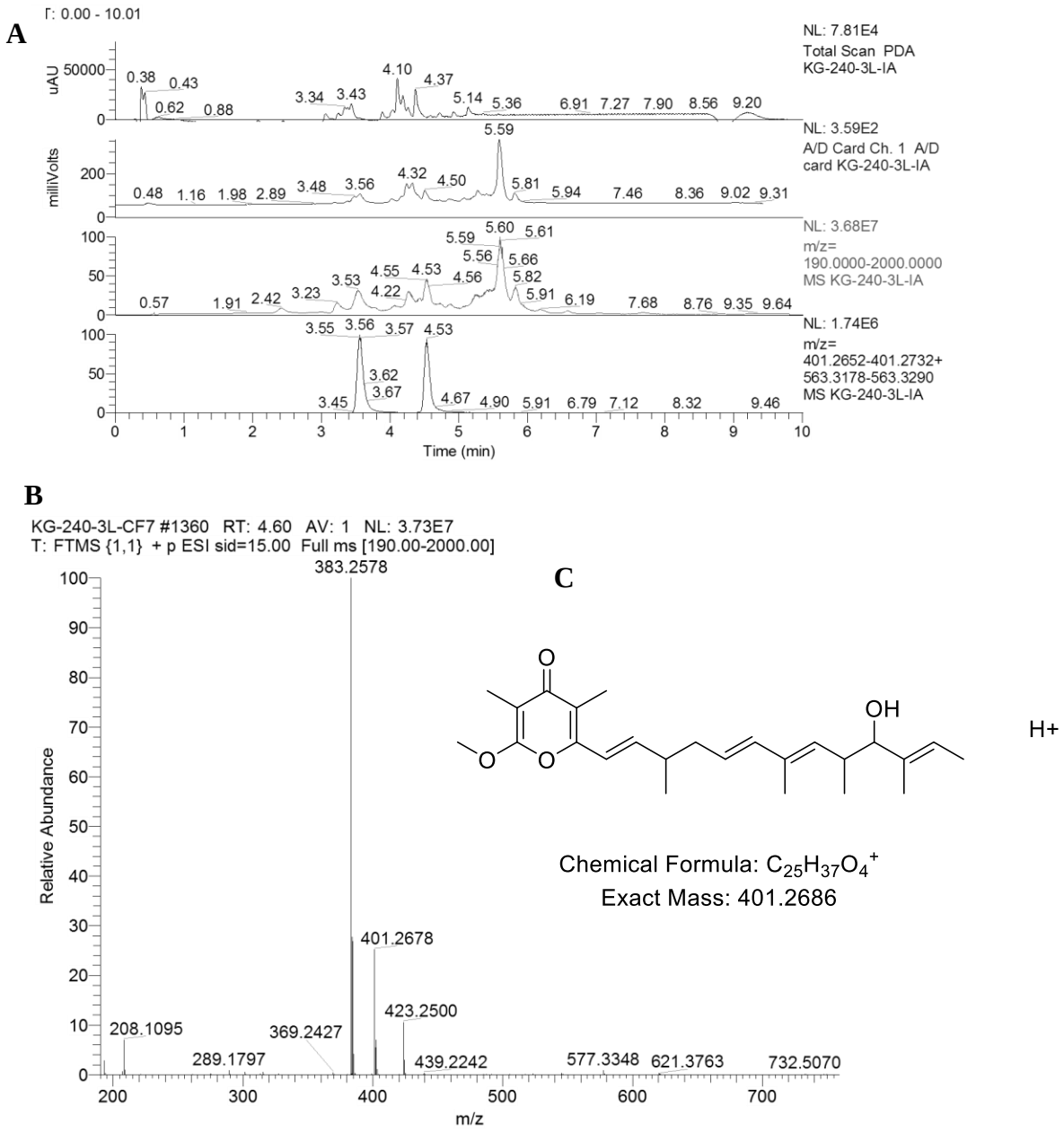
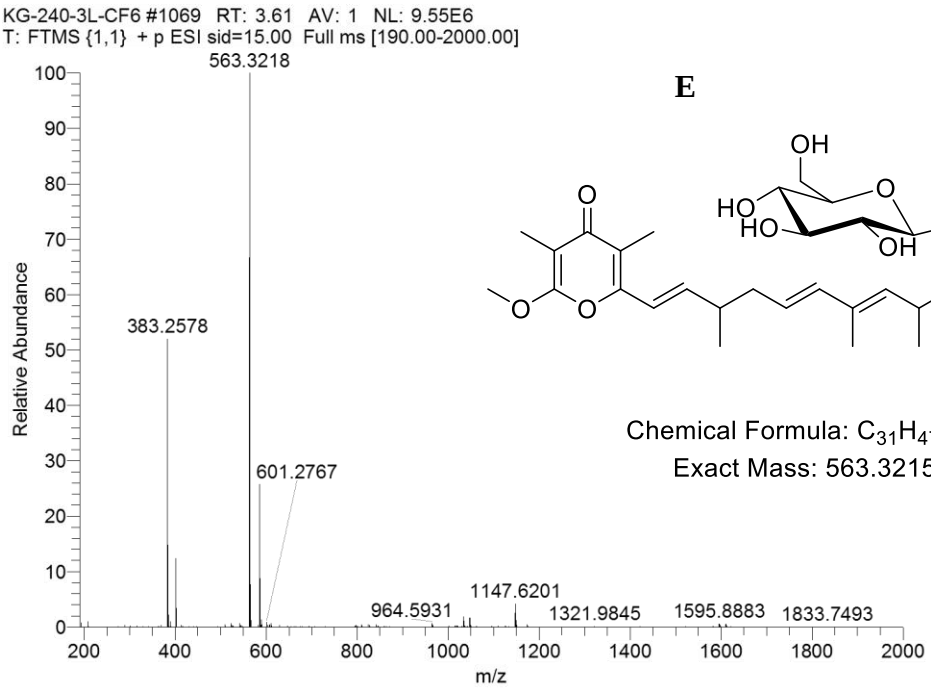


Figure 2.9. **A**. Chromatograms of the crude extracts from RKKD-240 grown in ISP2. Three detectors were used to generate 3 chromatograms for the same sample: PDA, ELSD and +ESI-MS. The fourth chromatogram shows SIM of the minor metabolites RKKD240_401 and RKKD240_563. **B**. The high resolution mass spectrum of RKKD240_401 **C**. The molecular structure of the actinopyranone natural product, PM05046. **D**. The high resolution mass spectrum of RKKD240_563. **E**. The molecular structure of PM050511.

D



E

Chemical Formula: $C_{31}H_{47}O_9^+$
Exact Mass: 563.3215

Figure 2.8. Continued from previous page.

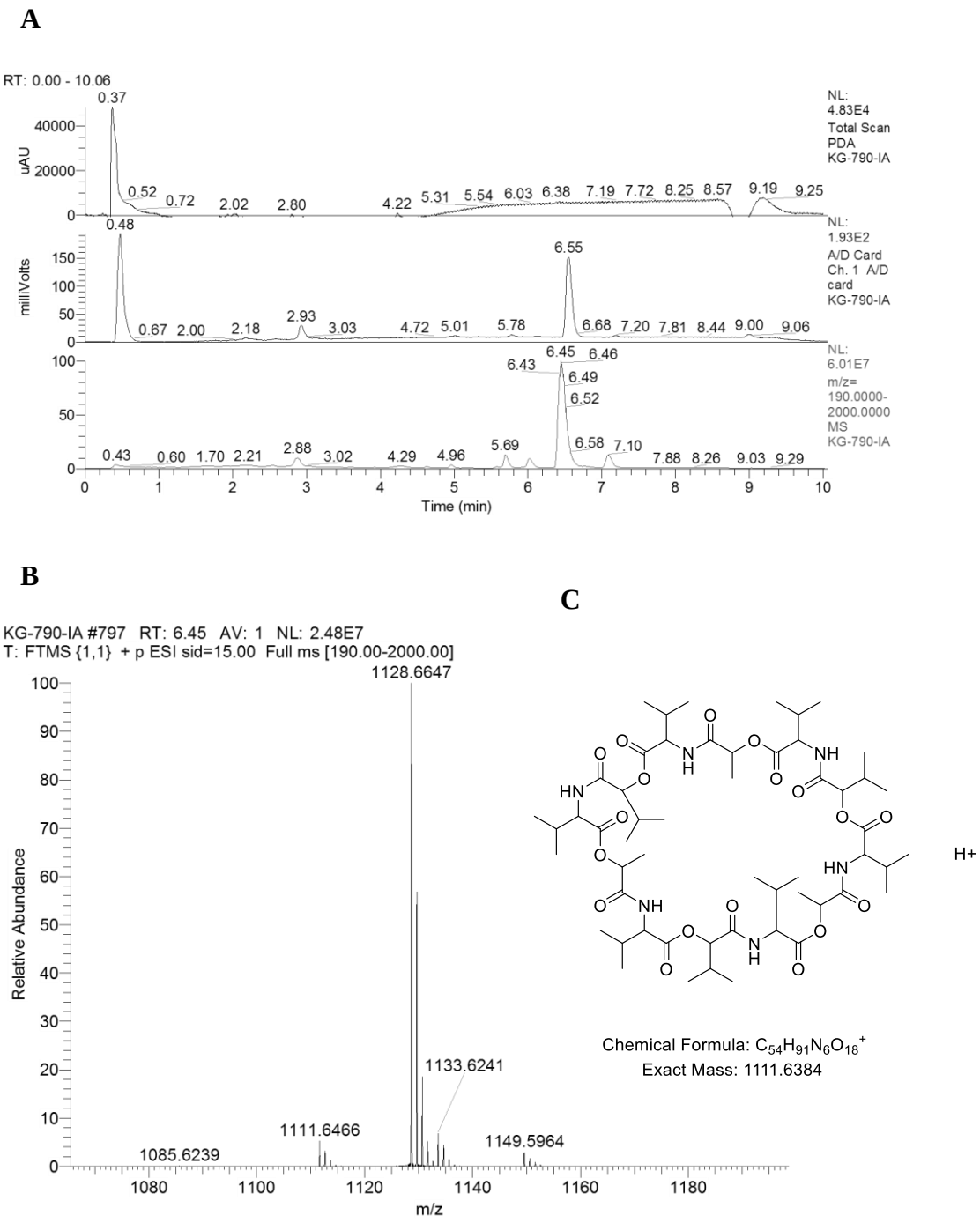
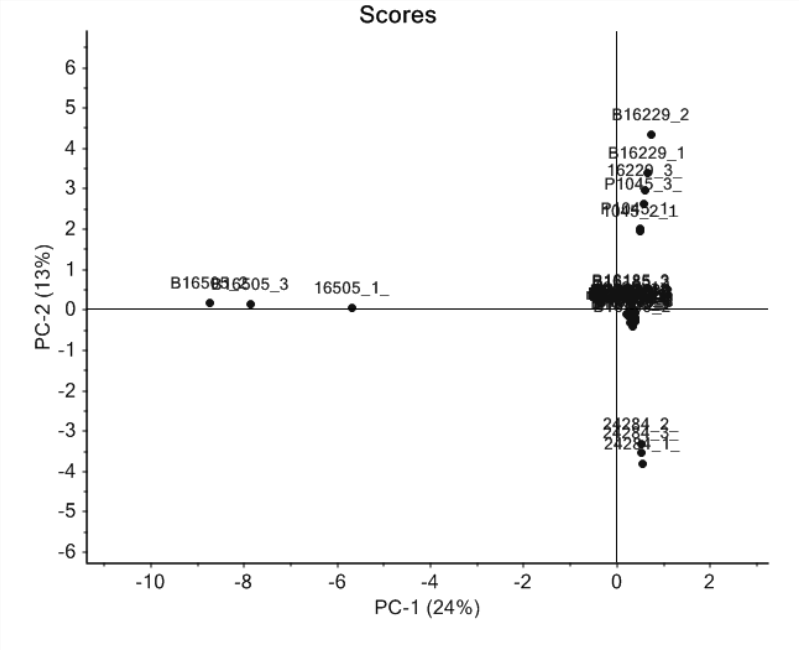


Figure 2.10. **A**. Chromatograms of the crude extracts from RKKD-790 grown in ISP2. Three detectors were used to generate 3 chromatograms for the same sample: PDA, ELSD and +ESI-MS. **B**. The high resolution mass spectrum of RKKD790_1128 **C**. The molecular structure of valinomycin.

2.3.4 Multivariate analysis of small scale fermentations: *Kitasatospora* library

Twelve strains of *Kitasatospora* were investigated for their metabolite production using the LCMS/PCA based chemical profile screen. All twelve strains were fermented in triplicate, extracted and analyzed by LC-HRMS. To ensure continuity, methanol blanks were used at the beginning, end and internally for each LC-HRMS sequence, and a retention time standard sample containing reserpine was injected at the beginning of each sequence. Raw LC-MS profiles were grouped and processed and then visualized using principal component analysis. A threshold of 1E4 was used during peak detection in order to ensure that only metabolites and not artifacts were detected and contributed to prioritization of strains for the isolation of metabolites. A total of 642 distinct mass features were detected after data processing with MZmine 2. Of those features, 190 were detected in MeOH blanks, media blanks, or internal reserpine blanks and were removed from the features list. Peak intensities were normalized to a value of one for all detected features to remove intensity bias associated with highly abundant metabolites. Principle component analysis was then used to visualize the processed data matrixes using The Unscrambler. Results from fermentations in ISP2 medium and lean production medium (1045) were visualized using separate principal component analyses. Scores plots of PC1 vs. PC2 to PC3 vs. PC4 were used to analyze the metabolomes of isolates from the *Kitasatospora* library, which are shown in Figures 2.11 and 2.12.

A



B

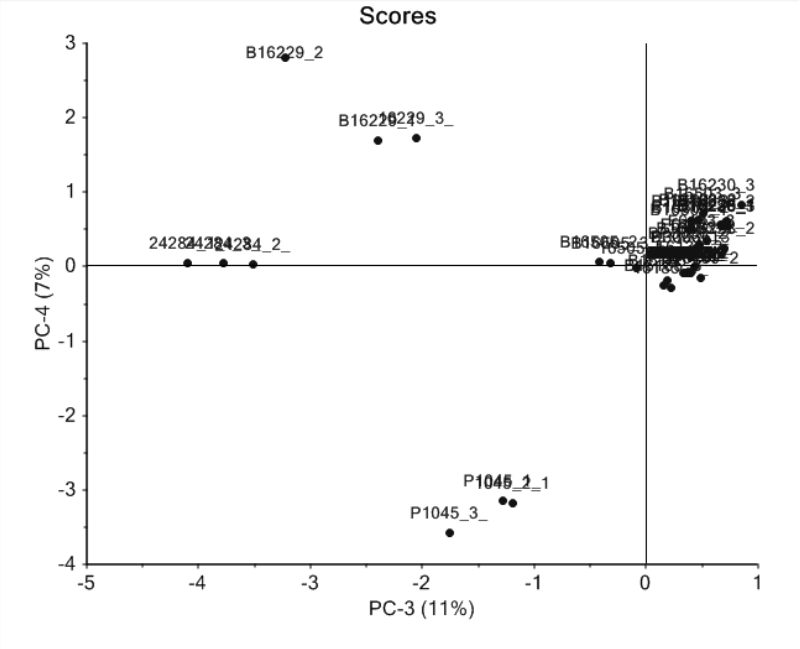


Figure 2.11. Principal component analysis scores plots of *Kitasatospora* library grown in 1045 medium.

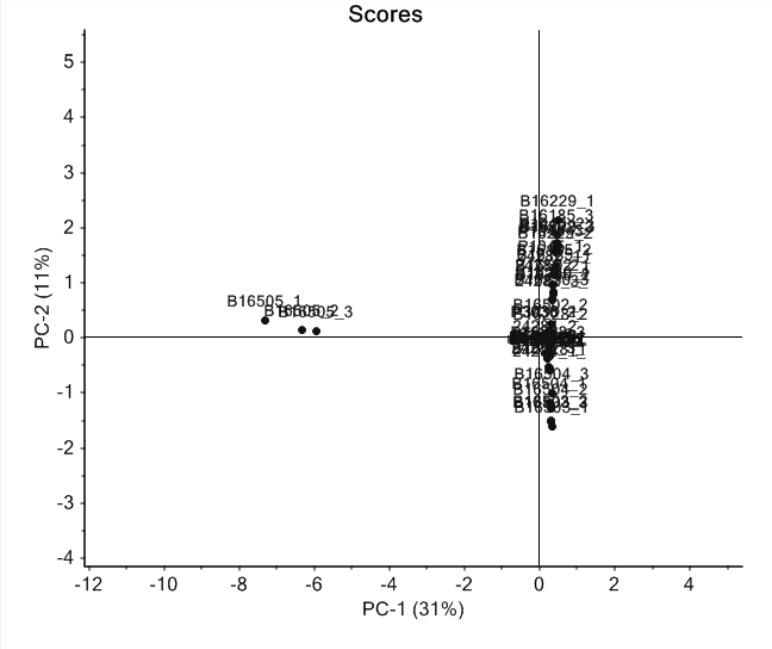


Figure 2.12. Principal component cnalysis ccores clots of PC1 vs PC2 of *Kitasatospora* library grown in ISP2 Medium.

Based on the scores plots, four strains were identified as producers of unique metabolites and selected as hits from the LC-MS/PCA screen. The following hits were identified: *Kitasatospora cystarginea* NRRL B-16505 in 1045 and ISP2 from PC1 vs. PC2; *K. kifunensis* NRRL B-24284, *K. griseola* NRRL B-16229 and *K. griseola* MF730-N6 BP-1045 in 1045 from PC3 vs. PC4. These strains were considered hits since they clustered away from all of the other strains in the corresponding scores plots, and all three replicate fermentations clustered together indicating consistent production of metabolites in replicate fermentations. No other scores plots for analyses of either fermentation medium showed strains that produced unique and consistent chemical profiles, as visualized using PCA. The strain *K. griseola* MF730-N6 BP-1045 was deprioritized because at the time it was under investigation by another member of the Kerr lab. Upon further investigation, the novel natural product satosporin was identified by Jennifer C. Arens from this strain[76].

2.3.5 Identification of metabolites: *Kitasatospora* library

The three remaining strains from the *Kitasatospora* library that were identified as hits from the LC-MS/PCA chemical screening method were fermented in larger volumes. *K. cystarginea* NRRL B-16505 was selected because of its unique metabolite profile when fermented in both ISP2 and lean production medium 1045. Metabolites were produced at a higher abundance when fermented in 1045 medium, therefore 1045 medium was used for all subsequent fermentation scale ups. The initial small scale fermentation resulted in the production of three major metabolites: 16505_357, 16505_371 and 16505_954. These metabolites did not correspond to any known compounds and were subsequently identified as novel natural products. The full

characterization of these compounds are discussed in Chapters 4 and 5. Several large-scale fermentations of this strain were performed, which resulted in extensive analysis of the metabolite production (Figure 2.13). In addition to the novel metabolites that were identified, this strain also produced several known compounds. The metabolite with an exact mass of *m/z* 211.0868 $[M+H]^+$ corresponded to the known compound 1-acetyl-β-carboline within 0.95 ppm of the predicted mass[77]. The identity of this metabolite was confirmed using NMR (Appendix, Figure 8.4). A minor metabolite that was detected in fractions of subsequent large-scale fermentations of this strain had an exact mass of *m/z* 490.2848 $[M+H]^+$. Database searches revealed that this exact mass was the same at the known compound migrastatin within 1.8 ppm[78]. NMR spectra of fractions containing this metabolite confirmed that the structure was the same as migrastatin (Appendix, Figure 8.5). Lastly, several members of the fattiviracin family of natural products were produced by *K. cystarginea* B-16505. Fattiviracin FV-10 and FV-13 as well as a fattiviracin analog with an exact mass of *m/z* 1529.9984 [M+H]+ were identified using HRMS and MS/MS and NMR (Appendix, Figure 8.8)[79].

The strain *K. kifunensis* NRRL B-24284 was identified as a hit from the LC-MS/PCA chemical screening method. Analysis of the small scale fermentation extract using LC-MS revealed that this strain produced a very complex mixture of metabolites making it difficult to dereplicate using databases (Figure 2.14). After fractionation of the crude extract, one fraction primarily contained a single metabolite with an exact mass of *m/z* 720.2941$[M+H]^+$. This exact mass corresponded to naphthomycin when queried in AntiBase within 0.3 ppm. The UV absorbance, and NMR spectra were used to confirm that the structure was naphthomycin (Appendix, Figure 8.7)[80].

K. griseola NRRL B-16229 was investigated for the production of novel metabolites. When fermented on a small scale, one major and one minor metabolite were produced. The exact masses of these metabolites did not correspond to any known natural products and therefore this strain was prioritized. The exact masses of these putatively novel metabolites were *m/z* 542.3230 $[M+H]^+$ and *m/z* 556.3386 $[M+H]^+$. All subsequent fermentations of this strain produced either none or a very low abundance of these two metabolites and the identities of these compounds were not determined.

K. mediocidica NRRL B16109 was not identified as a hit from the LC-MS/PCA screen however based on the raw LC-MS profile of the crude extract it produced a relatively simple mixture of metabolites when fermented in ISP2. The most abundant metabolite had an exact mass of *m/z* 686.4716 $[H+H]^+$ (Figure 2.15). This exact mass corresponded to the hexapeptide, pepstatin A within 1.7 ppm compared to the predicted mass[81]. NMR was used to confirm that the identity of 16109_686 was pepstatin A. Pepstatin A is a commonly produced metabolite that has been isolated from several different actinomycetes[81] (Appendix, Figure 8.8).

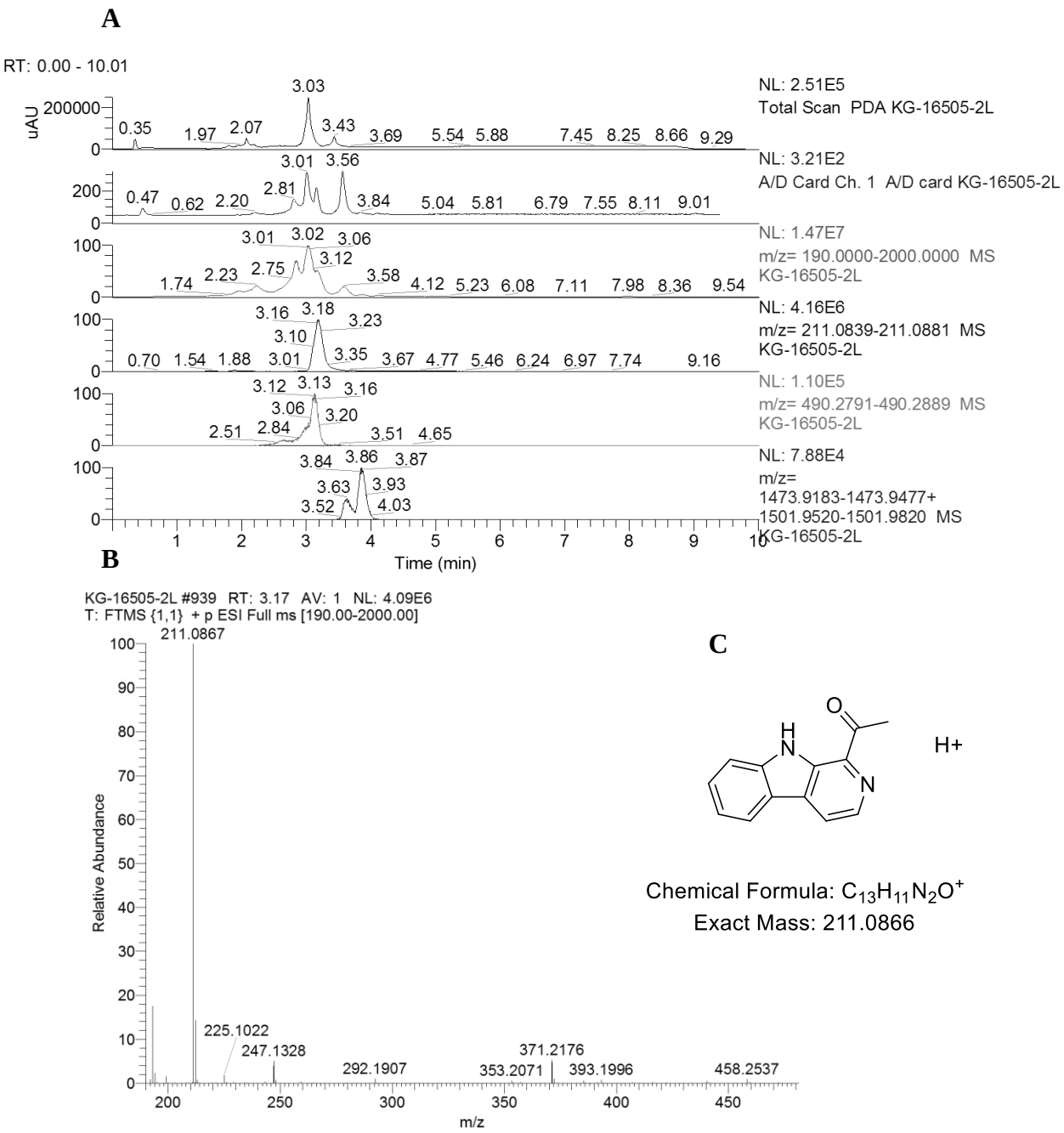


Figure 2.13. **A**. Chromatograms of the crude extracts from K. cystarginea NRRL B-16505 grown in 1045.
Three detectors were used to generate 3 chromatograms for the same sample: PDA, ELSD and +ESI-MS. The fourth, fifth and sixth chromatograms show SIM of the minor metabolites 16505_211, 16505_490 and 16505_1501 respectively. **B**. The high resolution mass spectrum of 16505_211. **C**. The molecular structure of 1-acetyl-β-carboline. **D**. The high resolution mass spectrum of 16505_490. **E**. The molecular structure of migrastatin. **F**. The high resolution mass spectrum of 16505_1501. **G**. The molecular structure of fattiviracin FV13.

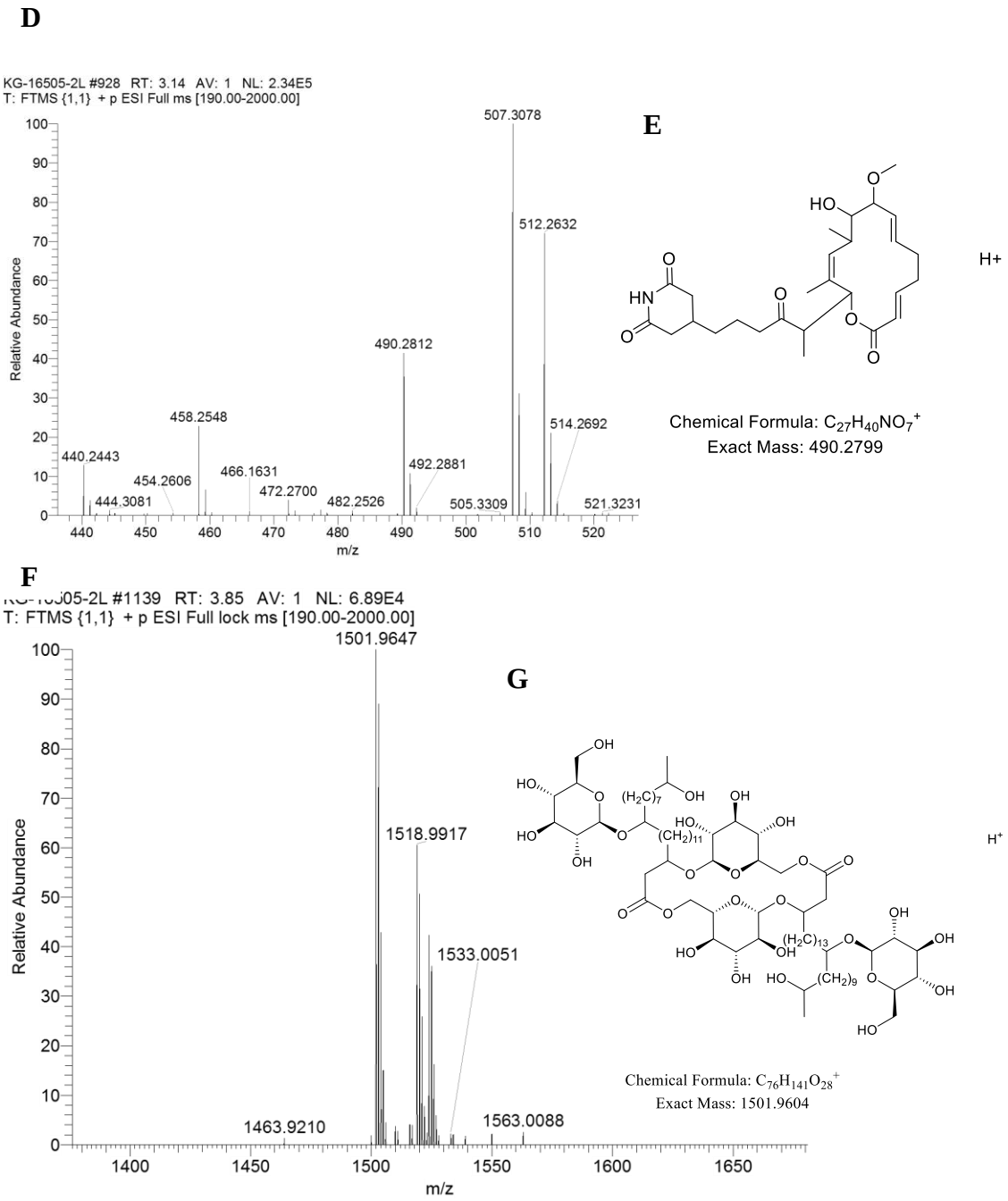


Figure 2.13. Continued from previous page.

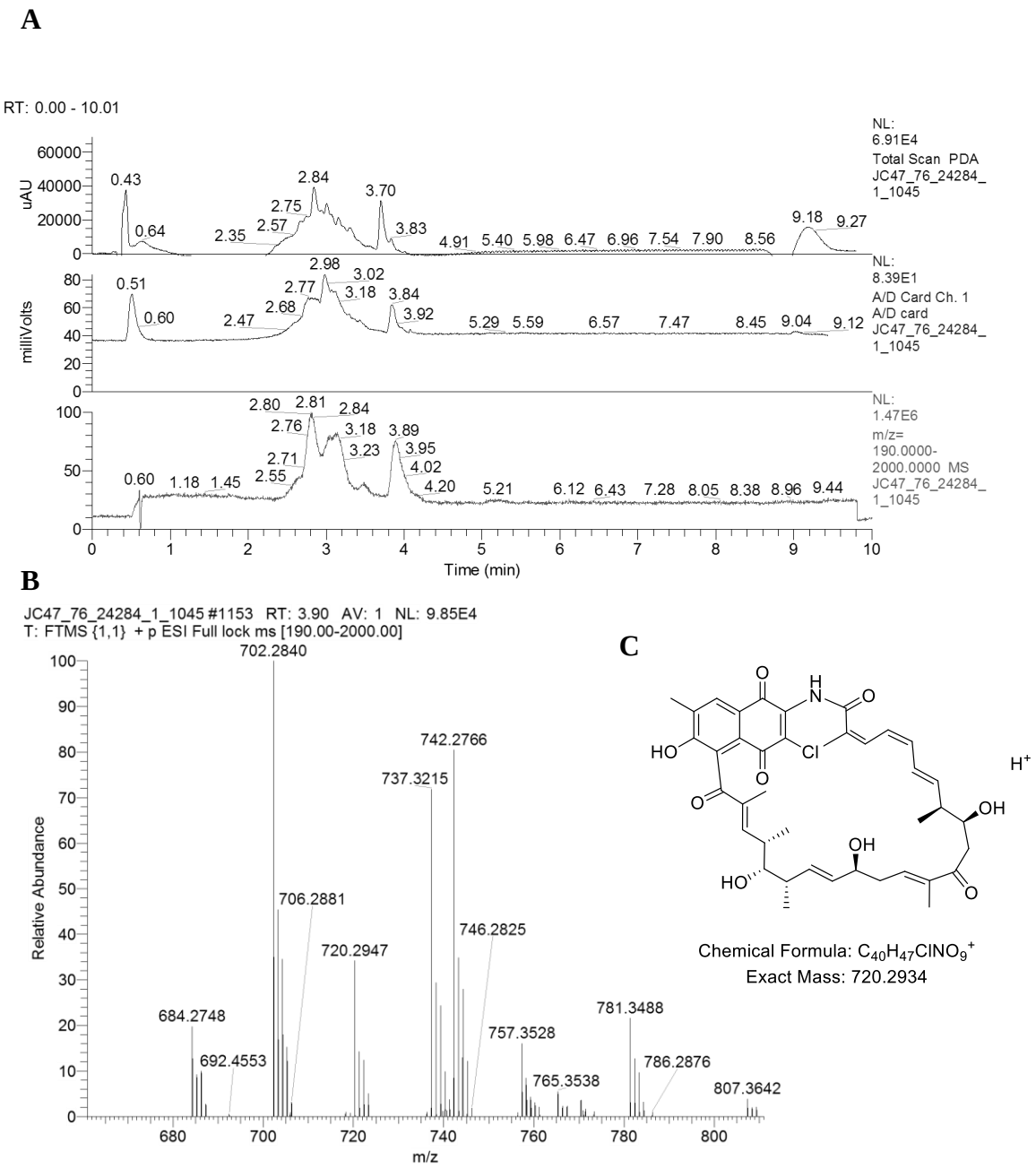


Figure 2.14. **A**. Chromatograms of the crude extracts from *K. kifunensis* NRRL B-24284 grown in 1045.
Three detectors were used to generate 3 chromatograms for the same sample: PDA, ELSD and +ESI-MS. **B**. The high resolution mass spectrum of 24284_720. **C**. The molecular structure of naphthomycin A.

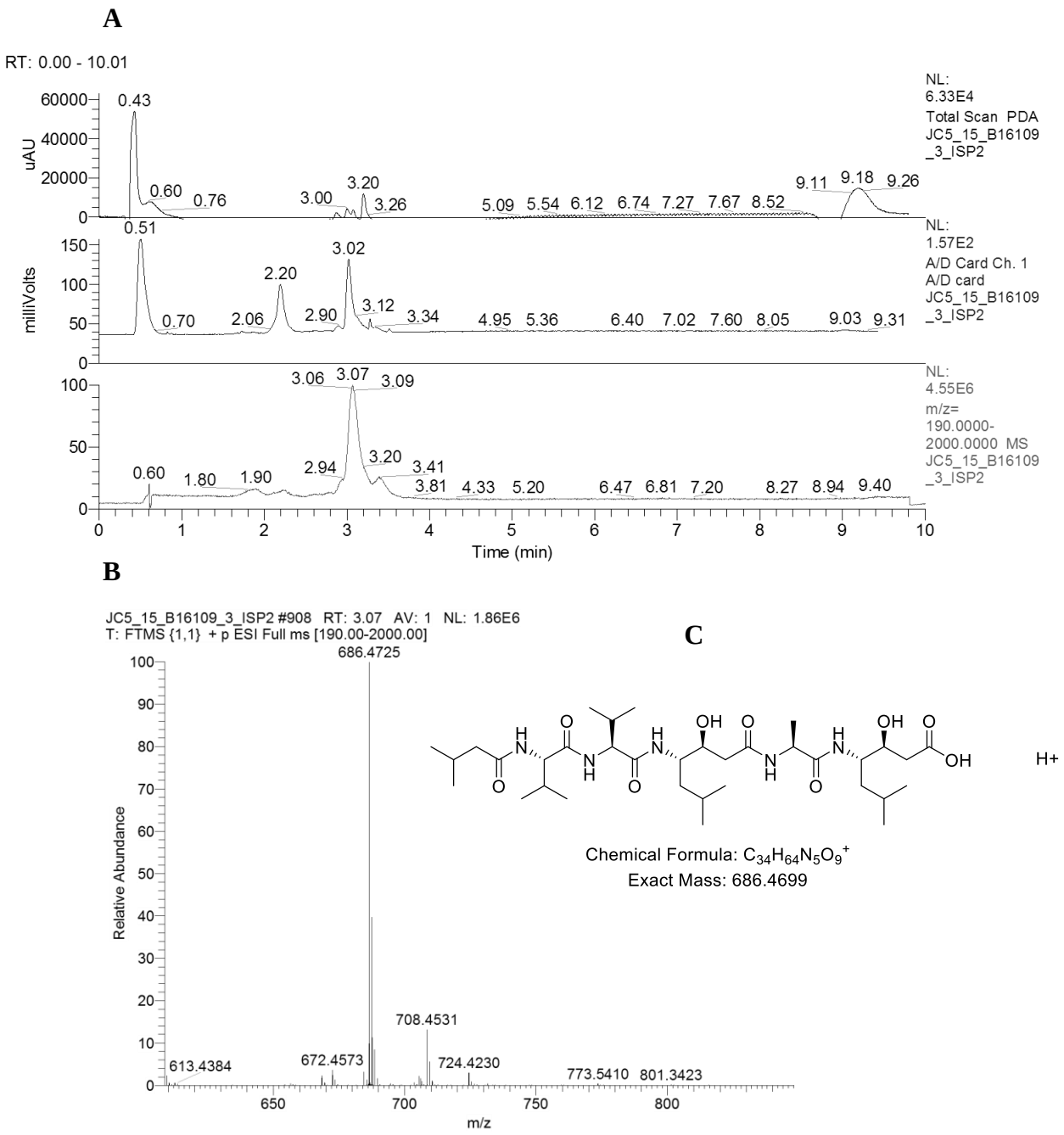


Figure 2.15. **A**. Chromatograms of the crude extracts from K. mediocidica NRRL B-16109 grown in ISP2.
Three detectors were used to generate 3 chromatograms for the same sample: PDA, ELSD and +ESI-MS. **B**. The high resolution mass spectrum of 16109_686. **C**. The molecular structure of pepstatin A.

2.3.6 Comparison of screening methods

The approach used for this natural product discovery program involved the combination of antimicrobial assays and LC-MS/PCA based chemical screening for the rapid identification of novel, bioactive natural products. When used alone for the identification of bioactive metabolites from *Streptomyces*, bioassay driven strain prioritization often returns primarily known natural products that have already been investigated[35]. The results observed in this study for the RKKD library followed this same trend. Six *Streptomyces* strains were selected and five were investigated further for production of bioactive natural products. Known compounds were identified from all of these strains. The strain RKKD-790 highlights the efficiency of LC-MS/PCA to identify putatively new compounds and known compounds. When bioactivity was used to isolate active metabolites from this strain, a known compound was identified (valinomycin). However, RKKD-790 also produced a novel lasso peptide that was not active in the bioassays and was instead identified as putatively novel from the LC-MS/PCA screen. If bioactivity alone was used, the known antimicrobial compound valinomycin would have been the only metabolite identified from this strain, and the new lasso peptide would have been missed. This highlights the importance of using a chemical screening approach, or a combination of screening methods in order to identify new metabolites.

The usefulness of LC-MS/PCA not only allows for identification of metabolites that are not active in specific bioassays, but it also allows for much more rapid identification of known compounds, and therefore efficient strain selection for fermentation scale ups. The strain RKKD-760 was prioritized because of its unique metabolite profile that was identified by LC-MS/PCA. The metabolites produced by this

strain were very efficiently identified as members of the arylomycin family of natural products from a single small scale fermentation. No additional fermentations of this strain were required to characterize these metabolites, which highlights how using LC-MS/PCA streamlines the process of novel metabolite discovery. This method allows for more efficient and informed strain selection for downstream characterization which are often very time consuming.

Based on the results from the RKKD library, it was observed that the LC-MS/PCA screening approach was efficient at identifying putative identities of natural products and was a useful method to prioritize and select strains for larger scale investigations. In contrast, bioactivity screening alone did not result in the identification of any novel natural products from this bacterial library. Based on these results, the *Kitasatospora* library was screened using LC-MS/PCA alone.

The effectiveness of LC-MS/PCA at identifying stains with unique metabolomes is also highlighted for the strain *K. cystarginea* NRRL B-16505. The chemical screening approach showed that this strain had a significantly different chemical profile than the rest of the *Kitasatospora* strains that were analyzed. Upon further investigation, it was identified that this strain produced a novel lipopeptide and two novel β-lactone containing natural products. After characterization these novel natural products they were evaluated for antimicrobial activity. None of these novel compounds possessed antimicrobial activity against the pathogens showing that if bioassay-based prioritization using antimicrobial assays was used to screen the *Kitasatospora* library, then this strain would have been overlooked and these novel compounds would not have been discovered. A chemical screening approach like LC-MS/PCA, a high-content bioassay or a combination

of both are the most efficient and effective way to analyze natural product libraries for drug discovery programs in the 21st century[20].

2.4 CONCLUSIONS

Two libraries of actinomycetes were investigated for the production of novel natural products using an LC-MS/PCA metabolomics screening approach. This method was utilized to efficiently prioritize strains for downstream characterization based on the production of unique metabolites among a library of similar bacterial isolates. A library consisting primarily of *Streptomyces* spp. was selected based on a number of criteria to investigate for novel natural products. Members of the RKKD library were cultured from temperate marine sediment in Newfoundland and New Brunswick, Canada[47,48]. The areas where these isolates were collected represent unique biological niches, which correlates to unique biological diversity of the microbiome[47,48]. Members of this library represented an opportunity to investigate for novel metabolites since greater biological diversity directly correlates to greater chemical diversity[15]. Investigation of this library resulted in identification of a novel lasso peptide at the time of isolation, using only a single culture medium. Given the use of a single culture medium, and the extensive investigation of *Streptomcyes* spp. for natural product production in the literature, identification of this novel lasso peptide from this *Streptomyces* library was an exciting discovery. This highlights the effectiveness of the LC-MS/PCA screening approach to identify the production of novel metabolites among libraries that produce several known compounds, and is useful for dereplication of such libraries.

A library consisting of *Kitasatospora* spp. was also investigated for the production of novel natural products using LC-MS/PCA. . Genome sequencing has

revealed that *Kitasatospora* possess a similar number of biosynthetic gene clusters as their close relative *Streptomyces.* However, this genus has not been investigated to the great extent that *Streptomyces* has for drug discovery programs, making them a useful resource for discovering novel metabolites. Two novel compounds were identified from this small library, which shows that members of the genus *Kitasatospora* represent a unique, and underexplored resource for the discovery of novel metabolites.

LC-MS/PCA was used to screen both bacterial libraries, and antimicrobial assays was used for one of the libraries to prioritize strains and isolate bioactive metabolites. In contrast to bioassay screening, which only lead to the isolation of known natural products, LC-MS/PCA helped to highlight the production of several novel metabolites. A new lasso peptide was identified from the fermentation extracts of *Streptomyces* sp. RKKD-790, and a novel lipopeptide and two β-lactone containing peptides were isolated from *K. cystarginea* NRRL B-16505.

Using LC-MS/PCA to analyze and visualize the variance in metabolite profiles is a very useful tool to rapidly dereplicate bacterial libraries and prioritize strains for additional investigation[21]. The time consuming process of scaling up fermentations, purification of metabolites, and structural characterization by analysis of spectroscopic data makes selection of strains an essential step for natural product programs[22]. LC-MS/PCA allows for the identification of unique metabolomes and dereplication against known natural products through the use of high resolution mass spectrometric analysis and comprehensive natural product databases[21,22]. Using LC-MS/PCA to screen microbial libraries for the production of unique metabolites greatly improves the efficiency of natural product discovery programs.

CHAPTER 3: ISOLATION AND CHARACTERIZATION OF A TYPE II LASSO PEPTIDE PRODUCED BY *STREPTOMYCES* SP. RKKD-790

3.1 INTRODUCTION

The identification of new natural products has long been used for the development of pharmaceutical agents. In the 21st Century, a new major class of natural products has been identified and named ribosomally-synthesized and post-translationally modified peptides (RiPPs)[82,83]. As the name suggests, RiPPs are peptide natural products that, unlike other peptides that are synthesized by non-ribosomal peptide synthetase enzymes, are synthesized directly through translation by the ribosome. They are then modified to varying degrees by tailoring enzymes to generate the final natural product structure. RiPPS have become a very large and structurally diverse class of natural products that has rapidly grown due to advances in genome sequencing efforts. Several members of this natural product class have very potent biological activities, with the most commonly observed application being antibacterial activity.

Lasso peptides are a group of RiPP natural products characterized by their unique knotted structure[84]. The lasso fold is formed by threading of the C-terminal linear tail through a macrolactam ring formed between the amine group of either glycine, alanine, serine or cysteine residue at position 1, and the side chain carboxylic acid group of either glutamic acid or aspartic acid at position 7, 8 or 9[85]. There are three classes of lasso peptides that are characterized by the presence of disulfide bridges between cysteine residues in which class I contains two[86,87,88], class II contains zero [89,90,91,92,93], and class III contains one[94,95] intramolecular disulfide bridge as shown in Figure 3.1. The disulfide bridges assist in stabilizing the threaded structures of class I and III, while class II are stabilized by bulky side chains in the loop region directly above the ring, and the tail

region below the ring that prevent unthreading[84,85]. Synthetic approaches to obtain the knotted structure of lasso peptides have been unsuccessful to date[96].

Lasso peptides have been found to have a wide variety of biological activities[84,85]. Many act as receptor antagonists as well as possess antimicrobial properties against bacteria closely related to their producing bacterium. Peptides generally have high specificity for their molecular targets and low toxicity; however their use as clinical agents is limited due to poor pharmacokinetic properties[97, 98.] Since the threaded structures of the lasso peptides cause them to be extremely stable to heat, chemical and proteolytic degradation, they may be useful for overcoming the rapid degradation observed by less stable peptides[99]. The lasso peptide microcin J25 (MccJ25) has been shown to assist in overcoming these limitations by acting as an effective molecular scaffold. Integrin proteins have been shown to play a role in angiogenesis and tumor growth. Integrins interact with a specific peptide sequence Arg-Gly-Asp (RGD), named the integrin binding motif. The RGD integrin binding motif was grafted onto the loop region of MccJ25, which resulted in a highly stable and nontoxic peptide with high binding affinity towards $\alpha_v\beta_3$ integrin[99]. This is an excellent example of the usefulness of this unique class of natural products as pharmaceutical agents.

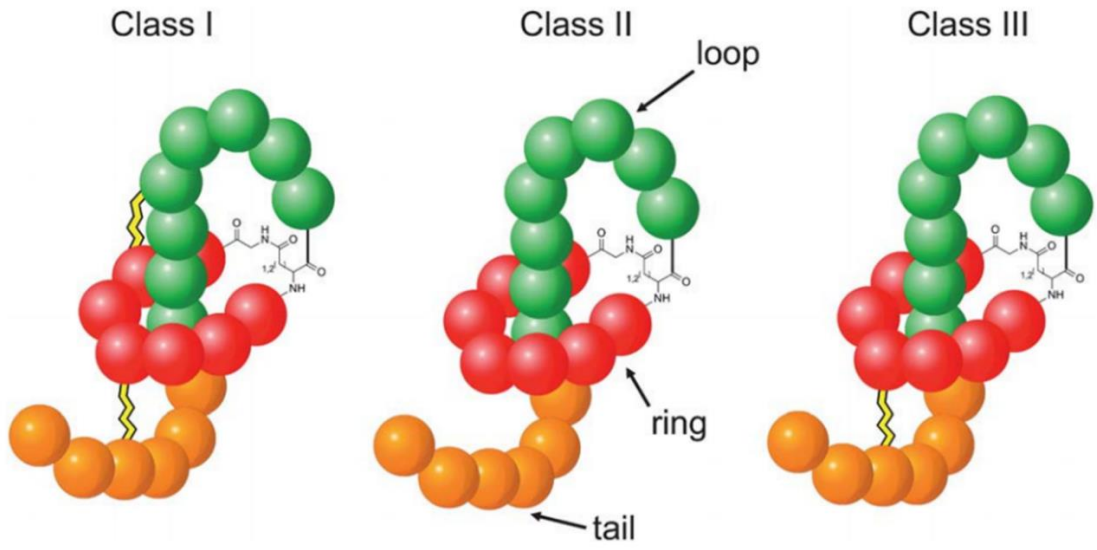


Figure 3.1 General conformations of the three classes of lasso peptide natural products[84].

By utilizing a metabolomics screening approach involving principal component analysis (PCA) described in Chapter 2, the fermentation extract of *Streptomyces* sp. RKKD-790 was identified as unique based on the metabolite profile. *Streptomyces* sp. RKKD-790 produced the secondary metabolite referred to as KG790_1592, which was responsible for the unique profile. Herein the isolation, structure elucidation, 3D molecular modeling and biological activity assessment of KG790_1592 is reported.

3.2 MATERIALS AND METHODS

3.2.1 Fermentation, extraction and purification of KG790_1592

Streptomyces sp. RKKD-790 was cultured from marine sediment collected from Bonne Bay, Newfoundland, Canada[48]. *Streptomyces* sp. RKKD-790 was streaked onto an ISP3[49] agar plate and grown at 30 ºC for 5 days. A single colony was used to inoculate 50 mL x 3 of ISP2[49] broth in 250 mL flasks and cultivated at 30ºC with shaking at 200 rpm for 3 days. In 2.8 L fernbach flasks, 4 L of fresh ISP2 broth (1 L x 4) was inoculated with 250 mL of the seed culture and cultivated at 30ºC with shaking at 200 rpm for 7 days. Fermentation broth was extracted by partitioning with 4 L of ethyl acetate and dried by evaporating under reduced pressure. The extract was dissolved in 500 mL of 8:2 acetonitrile:diH_2O and partitioned with 500 mL of hexanes.

The aqueous acetonitrile fraction was separated by reverse phase flash chromatography using a CombiFlash Rf (Teledyne ISCO) with a 15.5 g C_{18} column (High Performance GOLD, RediSep Rf), and was eluted with diH_2O (solvent A1) and MeOH (solvent B1). A linear gradient was used, increasing from 10 % to 100 % solvent B1 over 13 minutes followed by a 5 minute wash with 100 % solvent B1. A flow rate of

30 mL/min was used and eluent was detected by UV (214 nm). The desired metabolite, KG790_1592 eluted at 12.5 minutes.

KG790_1592 was purified using a Thermo HPLC system with a semi-preparative Gemini 110 Å C_{18} 5µm 250x10.00 mm column (Phenomenex) using diH_2O/0.1 % formic acid (solvent A2) and MeOH/0.1 % formic acid (solvent B2). KG790_1592 was eluted with a linear gradient increasing from 50 % to 80 % solvent B2 over 17 minutes, followed by a 100 % solvent B2 wash for 14 minutes. KG790_1592 eluted at 14.9 minutes and was detected by ELSD and UV at 220 nm and 280 nm.

LC-MS was used to monitor fractions for the presence of KG790_1592 by detecting *m/z* 1592.8079 using the standard Kerr lab "Chemical Screening" LC-MS method. This method uses a Thermo Q Exactive HPLC system with an analytical Core Shell 100 Å C_{18} 1.7 µm 50 x 2.1 mm column (Kinetex) with diH_2O/0.1 % formic acid (solvent A3) and acetonitrile/0.1 % formic acid (solvent B3). A linear solvent gradient from 5 to 100 % solvent B3 over 4.8 minutes followed by a 3.2 minute wash with solvent B3 was used with a flow rate of 500 µL/min. Eluent was detected by ESI-MS monitoring m/z 190-2000 in positive mode, ELSD, and PDA 200-600 nm.

Additional 5 L and 6 L fermentations of RKKD-790 were also cultivated, extracted and purified using the same methods described above in order to obtain enough pure KG790_1592 to perform full characterization and evaluation of biological activity.

3.2.2 Structure elucidation of KG790_1592

NMR spectra were obtained on a 600 MHz Bruker Avance III spectrometer equipped with a 1.7 mm inverse probe. For structure assignments, ^{1}H, COSY, TOCSY,

ROESY, HSQC, HMBC and NOESY experiments were performed, and analyzed using Bruker TOPSPIN 2.1. Chemical shifts (δ) are reported in ppm and were referenced to the DMSO-d_6 residual peaks at δ_H 2.50 ppm and δ_C 39.51 ppm and coupling constants (J) are reported in Hz. NOE cross-peak volumes were integrated manually using Bruker TOPSPIN 2.1 to obtain distance constraints.

High-resolution mass spectra were measured on a Thermo LTQ Orbitrap Velos mass spectrometer. Tandem MS spectra were acquired by direct infusion of KG790_1592 at a rate of 2 µL/min using an ESI source and fragmented by higher energy collisional dissociation (HCD) at 29 eV.

KG790_1592: White powder; ^{1}H and ^{13}C NMR see Table 3.1; (+) HRESIMS *m/z* 1592.8077 $[M+H]^+$ (calculated for $C_{77}H_{110}N_{17}O_{20}$, 1592.8113, Δppm -2.16).

3.2.3 Stereochemical analysis by Marfey's Method.

Amino acid absolute configurations were determined by Marfey's Analysis of hydrolyzed KG790_1592[100]. 10µL of KG790_1592 (10 mg/mL in MeOH) was reacted with 200 µL of 6M HCl and heated to 120ºC for 24 hours. The reaction mixture was cooled and neutralized to pH 7 using 1N $NaHCO_3$. To the reaction vial, 10 µL of 1-fluoro-2,4-dinitrophenyl-L-alanine (L-FDAA) (10 mg/mL in acetone) was added followed by 50 µL of acetone and 100 µL of $NaHCO_3$. The reaction was heated to 80 ºC for 30 minutes before quenching with 500 µL of 1M HCl. The reaction mixture was analyzed by LC-MS and compared to authentic derivatized amino acid standards. Derivatized amino acids were analyzed by LC-MS using an LTQ Orbitrap Velos (ThermoScientific) HPLC system with Hypersil Gold 100Å 1.9µm C_{18} 50 x 2.1mm column (Thermo). A linear gradient from 95 % water/0.1 % formic acid (solvent A3) and

5 % acetonitrile/0.1 % formic acid (solvent B3) to 40 % solvent B3 over 55 minutes followed by a rapid increase to 100 % solvent B3 for 3 minutes was used with a flow rate of 400 µL/min. Eluent was detected by +ESI-MS monitoring m/z 120-800 and UV 200-600 nm. Racemization of tryptophan occurred under high reaction temperatures so reactions were repeated using milder conditions. 100 µg of KG790_1592 was reacted with 6M HCl at 80 ºC for 6 hours, followed by neutralization with 1N $NaHCO_3$. 10µL of L-FDAA (10 mg/mL in acetone) was added to the reaction vial followed by 50 µL of acetone and 100 µL of $NaHCO_3$ and heated to 40 ºC for 2 hours. The reaction was quenched with 500 µL of 1M HCl, and analyzed by LC-MS using the conditions described above.

3.2.4 3D Molecular modelling

To further elucidate the structural details of KG790_1592, the three-dimensional structure was modelled computationally. Six candidate structures were built and optimized using the Merck Molecular Force Field (MMFF) in the Spartan package (*Spartan '08*; Wavefunction, Inc). Experimental inter-hydrogen distances were calculated by integrating NOE cross-peak volumes manually using Bruker TOPSPIN 2.1. Experimentally determined inter-atomic distance constraints were then gradually implemented followed by refining and reoptimizing the resulting structures in a step-wise manner starting with the loop region, then the chain region, and finally H-H distances between the loop and chain region.

Once all experimental constraints were applied and the six candidate structures were built, all experimental constraints were relaxed and the structures were reoptimized using the more robust PM3 semi empirical method[101,102]. This resulted in a predicted

minimum energy geometry for each of the six candidate geometries. The experimental H-H distances could then be compared to the predicted structures for each geometry.

3.2.5 Stability assays

The stability of KG790_1592 was evaluated in the presence of purified carboxypeptidase Y. Carboxypeptidase Y (0.1 mg of , 95 U/mg, Sigma) was dissolved in 250 µL of 20 mM TRIS buffer (pH = 8.0). To the solution, 5 µL of KG790_1592 (10 mg/mL in DMSO) was added and incubated at 37 ºC. Aliquots of 50 µL were removed at the following intervals: T = 0h, 1h, 2h, and 3h. Once removed, the aliquots were diluted with 200 µL of 20 mM TRIS and incubated at 65 ºC for five minutes to inactivate the enzyme. The resulting solutions was analyzed by LC-MS using a Thermo Q Exactive HPLC system for the presence of intact KG790_1592 using the chemical screening conditions described above. The portion of intact RKKD790_1592 was calculated by comparing the peak area detected by +ESI-MS to the peak area at T=0 h.

The physiological stability of KG790_1592 was also evaluated in the presence of human serum. 50 µL of KG790_1592 (10 mg/mL in DMSO) was added to 500 µL of human serum type AB male (Sigma) and incubated at 37 ºC. 50 µL aliquots were removed at the following time (T) intervals: T= 0h, 1h, 2h, 3h, 7h, 12h, 24h, and 48h. At the time of removal, 200 µL of cold MeOH was added to precipitate large proteins and stop degradation. Reactions were centrifuged at 10,000 rpm for 15 minutes at 4 ºC, and supernatant was analyzed by LC-MS using a Thermo Q Exactive HPLC system for the presence of intact KG790_1592 using the chemical screening method described above. The percentage of intact RKKD790_1592 was calculated by comparing the peak area detected by +ESI-MS to the peak area at T=0 h.

3.2.6 Biological activity assays

Purified KG790_1592 was tested for antimicrobial activity against a panel of human microbial pathogens consisting of methicillin-resistant *Staphyloccocus aureus* ATCC 33591 (MRSA), vancomycin resistant *Enterococcus faecium* EF379 (VRE), *Staphylococcus warneri* ATCC 17917, *Pseudomonas aeruginosa* ATCC 14210, *Proteus vulgaris* ATCC 12454 and *Candida albicans* ATCC 14035. All assays were carried out according to Clinical Laboratory Standards Institute testing standards in a 96-well plate and in triplicate[53]. Control antibiotics used were vancomycin for MRSA and *S. warneri*, rifampicin for VRE, or nystatin for *C. albicans*. Cell growth was determined by measuring the optical density at 600 nm at time zero and after incubation for 22 hours at 37ºC using a Thermo Scientific Varioskan Flash plate reader. The change in OD_{600} after incubation was compared to untreated negative control wells and plotted as percent growth inhibition. KG790_1592 was dissolved in sterile 20 % DMSO and assayed at a final concentration of 50µg/mL.

KG790_1592 was also tested for antibiotic activity against *Arcanobacterium pyogenes* (AVC 3), *Corynebacterium pseudotuberculosis* (AVC 15), *Dermatophilus congolensis* (AVC 17), *Micrococcus luteus* (AVC 35), *Rhodococcus equi* (AVC 57) and *Mycobacterium smegmatis* in a disk diffusion assay. Assay conditions were performed in accordance with Kirby-Bauer disk diffusion susceptibility test protocol[103]. KG790_1592 was evaluated at a concentration of 10 µg/disk in MeOH. Penicillin G (2 µg), vancomycin (5 µg), rifampicin (2 µg) and gentimicin (10 µg) were used as positive controls and MeOH was used as a negative control. Zones of inhibition were measured with calipers after 24 h of growth.

KG790_1592 was evaluated for cytotoxicity against human breast adenocarcinoma cells (ATCC HTB-26) and human foreskin BJ fibroblast cells (ATCC CRL-2522). Breast adenocarcinoma cells were grown in Dulbecco's Modified Eagle's Medium/Nutrient Mixture F-12 Ham containing fetal bovine serum (10%), penicillin (100 µU/mL) and streptomycin (0.1 mg/mL) with 5% CO_2 at 37ºC. Human foreskin BJ fibroblast cells were cultured in Eagle's minimal essential medium containing fetal bovine serum (10%), penicillin (100 µU/mL) and streptomycin (0.1 mg/mL) with 5% CO_2 at 37ºC. Both cells lines were grown to 80 % confluency and diluted with their respective growth medium without antibiotics. To a microwell plate, 90 µL of BJ fibroblast cells were added at a concentration of 1 x 10^4 cells/well and 90 µL of HTB-26 at a concentration of 5 x 10^3 cells/well and incubated with 5 % CO_2 at 37ºC for 24 hours. After 24 hours cells were treated with 10 µL of KG790_1592 at final concentrations ranging from 1 µg/mL to 128 µg/mL in sterile DMSO. Blank wells containing only growth medium and DMSO were used as positive controls, untreated wells containing growth medium, DMSO and cells were used as negative controls, and wells treated with either doxorubicin (HTB-26) or zinc pyrithione (fibroblast BJ) were used as positive treated controls. All samples were repeated in triplicate. Microwell plates were incubated with 5 % CO_2 at 37ºC for 72 hours (HTB-26) or 24 hours (fibroblast BJ). After incubation, alamarBlue® was added at a final concentration of 10 % v/v. Fluorescence was measured at 560 excitation and 590 nm emission at time zero, and four hours after addition of alamar blue using a Thermo Scientific Varioskan Flash plate reader. The change in fluorescence was used to calculate the percent viability of each treatment compared to negative control wells. The concentration required to reduce the cell growth

by 50 % (GI_{50}), were calculated by Prism 6.0 (GraphPad Software) using a nonlinear regression dose-response, variable slope model based on triplicate measurements ± standard deviation.

3.3 RESULTS AND DISCUSSION

3.3.1 Identification of *Streptomyces* sp. RKKD-790

The 16S rRNA gene sequence was used to identify the producing bacterium encoded RKKD-790. A nucleotide BLAST of the 16S gene sequence (1549 bp) showed similarity to 54 *Streptomyces* strains with less than 2 base pair differences. Due to the similarity of 16S rRNA sequences of the genus *Streptomyces,* this technique is only capable of identification at the genus level, and not robust enough to differentiate species within the genus *Streptomyces*[104]. The bacterial strain responsible for production of KG790_1592 was therefore classified as *Streptomyces* sp. RKKD-790.

3.3.2 Large-scale fermentation, extraction and purification of KG790_1592

Through the use of principal component analysis, *Streptomyces* sp. RKKD-790 was found to have a unique chemical profile among the subset of the RKKD bacterial library that was analyzed. Database searches of the compound with a pseudomolecular ion, $[M+H]^+$ m/z= 1592.8077 returned no previously discovered natural products. This lead to the selection of *Streptomyces* sp. RKKD-790 for fermentation scale up. A larger scaled fermentation was undertaken, culture broths were extracted with ethyl acetate, and the resulting organic extract was fractionated by reverse phase liquid chromatography and further purified using RP-HPLC to yield 3.5 mg of pure KG790_1592.

3.3.3 Structure elucidation and absolute configuration of KG790_1592

The molecular formula of the protonated adduct was calculated to be $C_{77}H_{110}N_{17}O_{20}^{+}$ from the exact mass *m/z* 1592.8077 $[M+H]^{+}$ and 1D and 2D NMR were used for complete structure elucidation (Appendix, Figures 8.9-8.17). All chemical shifts, and key 2D NMR correlations used for structure elucidation are listed in Table 3.1. COSY and TOCSY spectra were used in combination for assignment of amino acid side chain proton signals, and the HSQC spectrum was used for assigning carbon signals with the exception of quaternary carbons. The ROESY spectrum was used for sequence determination based on correlations observed between amide protons and α-protons of adjacent amino acids. Pro-6 contains a tertiary amide and therefore lacks an amide proton. Pro-6 instead was assigned based on the HMBC correlation between Ile-7 NH and Pro-6 αCH, the ROESY correlation between Ile-7 NH and Pro-6 βCH_2', and between Pro-6 δCH_2, δCH_2' and Lys-5 αCH. Ile-7 lacked TOCSY correlations between NH and βCH, γCH_3, γCH_2 and δCH_3 and therefore signals were assigned using COSY and HMBC correlations between adjacent protons. TOCSY patterns were used to discern signals in regions of the spectra with overlapping signals. The two proton signals with chemical shifts of 4.53 ppm were determined to be two distinct signals based on the two different carbon signals found in the HSQC spectrum. They were assigned to Ile-7 αCH and Ser-13 αCH based on the TOCSY correlations to Ser-13 βCH_2 and βCH_2' at 3.66 and 3.36 ppm as well as the TOCSY correlations to Ile-7 NH. The overlapping signals in the aliphatic region of the ^{1}H spectrum (0.43-1.11 ppm) were assigned using TOCSY patterns from NH or αCH signals which usually appeared in less crowded regions of the

spectra. Strong NOEs between Gly-1 NH and Asp-8 βCH_2 signals revealed the internal macrolactam ring structure. The planar structure of KG790_1592 is shown in Figure 3.2.

The amino acid sequence and lasso structure were confirmed based on MS/MS fragmentation patterns (Appendix, 8.10). The expected oxonium (b) and ammonium (y) cations that form through cleavage of peptide bonds were observed in the MS/MS spectrum, and are summarized in Table 3.2[105]. The unique b+y ion complexes that are commonly observed for lasso peptides, occur when the loop region of the linear tail fragments, but the tail remains threaded through the ring resulting in both distinct fragments contributing to the detected mass despite no longer being covalently attached. This pattern is characteristic of the threaded structure of lasso peptides and is observed in microcin J25 and capistruin[92].

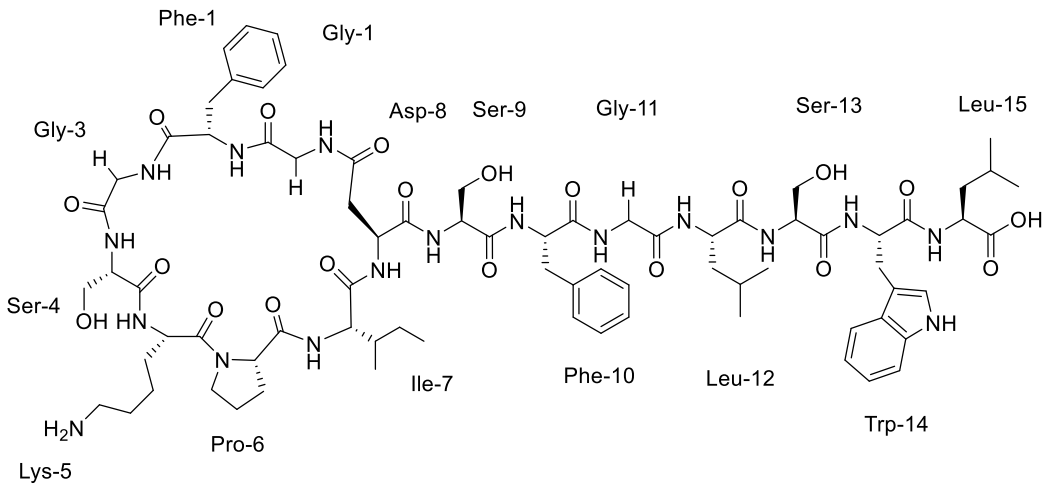


Figure 3.2. Molecular structure of RKKD790_1592.

Table 3.1. ^{1}H (600 MHz) and ^{13}C (150 MHz) and key 2D NMR data for RKKD790_1592 in DMSO-d_6.

Amino Acid	Position	$\delta^{1}H$ ppm	$\delta^{13}C$ ppm	COSY	ROESY	HMBC
Gly1	NH	7.70				
	αCH2	4.03	43.04	G1-NH	F2-NH	
	αCH2'	3.05	43.04	G1-NH, αCH2	F2-NH	
Phe2	NH	8.59				
	αCH	3.89	56.35	F2-NH	G3-NH	
	βCH2	3.40	33.44	F2-αCH		
	βCH2'	3.33	33.44	F2-αCH		
	Ar: H2,6	7.18	127.7	F2-Ar:H3,5		
	Ar:H3,5	7.30	128			
	Ar:H4	7.19	125.9	F2-Ar:H3,5		
Gly3	NH	7.87				
	αCH2	4.29	43.29	G3-NH	S4-NH	
	αCH2'	3.22	43.29	G3-αCH2	S4-NH	
Ser4	NH	7.43				
	αCH	4.43	55.49	S4-NH	K5-NH	
	βCH2	3.51	62.63	S4-OH, αCH, βCH2'		
	βCH2'	3.71	62.63	S4-OH, αCH		
	OH	4.89				
Lys5	NH	8.88				
	αCH	4.80	49.68	K5-NH	P6-δCH2, δCH2'	
	βCH2	1.54	31.23	K5-αCH, βCH2'		
	βCH2'	1.68	31.23	K5-αCH		
	γCH2	1.22	19.79	K5-βCH2, βCH2', δCH2		
	γCH2'	1.38	19.79	K5-βCH2, βCH2', δCH2		
	δCH2	1.30	25.82	K5-εCH2, εCH2', δCH2'		
	δCH2'	1.48	25.82	K5-εCH2, εCH2'		
	εCH2	2.53	37.51	K5-δCH2'		
	εCH2'	2.67	37.51	K5-δCH2'		
Pro6	αCH	4.30	56.72		I7-αCH	I7-NH
	βCH2	2.02	29.03	P6-αCH		
	βCH2'	1.92	29.03	P6-αCH, βCH2	I7-NH	
	γCH2	1.8	25.21	P6-βCH2, βCH2', δCH2, δCH2', γCH2'		
	γCH2'	2.16	25.21	P6-βCH2, βCH2', δCH2, δCH2', γCH2'		
	δCH2	3.62	46.98	P6-δCH2'		
	δCH2'	3.85	46.98			

Amino Acid	Position	$\delta^{1}H$ ppm	$\delta^{13}C$ ppm	COSY	ROESY	HMBC
Ile7	NH	6.58				
	αCH	4.53	58.88	I7-NH	D8-NH, P6-αCH	I7-γCH3
	βCH	2.04	40.4	I7-αCH		I7-γCH3, I7-δCH3
	γCH2	1.31	23.4	I7-βCH		I7-γCH3, I7-δCH3
	γCH2'	1.11	23.4	I7-βCH, γCH2		I7-γCH3, I7-δCH3
	γCH3	0.80	15.3	I7-βCH		
	δCH3	0.78	11.6	I7-γCH2, γCH2'		
Asp8	NH	8.44				
	αCH	4.73	48.6	D8-NH	S9-NH	
	βCH2	3.58	36.04	D8-αCH	S9-NH, G1-NH	
	βCH2'	2.72	36.04	D8-αCH, βCH2	G1-NH	
Ser9	NH	8.15				
	αCH	3.80	59.54	S9-NH	F10-NH	
	βCH2	3.25	60.15	S9-αCH, OH		
	βCH2'	3.40	60.15	S9-αCH, OH		
	OH	4.88				
Phe10	NH	8.11				
	αCH	4.61	52.89	F10-NH	G11-NH	
	βCH2	2.75	37.88	F10-αCH, βCH2'		
	βCH2'	3.10	37.88	F10-αCH		
	Ar: H2,6	7.04	128.7	F10-Ar:H3,5		
	Ar:H3,5	7.18	128.9			
	Ar:H4	7.09	125.9	F10-Ar:H3,5		
Gly11	NH	7.36				
	αCH2	3.47	42.68	G11-NH, αCH2'	L12-NH	
	αCH2'	4.11	42.68	G11-NH		
Leu12	NH	8.33				
	αCH	4.07	51.41	L12-NH	S13-NH	
	βCH2	0.73	37.4	L12-αCH, βCH2', γCH		L12-δCH3, L12-δCH3'
	βCH2'	1.77	37.4	L12-αCH		L12-δCH3, L12-δCH3'
	γCH	1.52	23.22	L12-βCH2, βCH2'		L12-δCH3, L12-δCH3'
	δCH3	0.80	15.35	L12-γCH		
	δCH3'	0.54	18.81	L12-γCH		L12-δCH3

Amino Acid	Position	δ^{1}H ppm	δ^{13}C ppm	COSY	ROESY	HMBC
Ser13	NH	9.31				
	αCH	4.53	53.76	S13-NH	W14-NH	
	βCH2	3.36	61.88	S13-αCH, βCH3'		
	βCH2'	3.66	61.88	S13-αCH		
	OH					
Trp14	NH	7.46				
	αCH	4.87	53.87	W14-NH	L15-NH	
	βCH2	3.05	25.21	W14-αCH, βCH2'		
	βCH2'	3.52	25.21	W14-αCH		
	Ar: NH	10.95				
	Ar:H1	7.87	124.5	W14-ArNH		
	Ar:H2	7.59	119.2	W14-ArH3		
	Ar:H3	6.92	117.9	W14-ArH2, ArH4		
	Ar:H4	7.02	120.4	W14-ArH5		W14-Ar:H2
	Ar:H5	7.29	110.5	W14-ArH4		
Leu15	NH	7.14				
	αCH	3.92	52.17	L15-NH		
	βCH2	0.85	42.32	L15-αCH		
	βCH2'	0.53	42.32	L15-αCH, βCH2		L15-δCH3, L15-δCH3'
	γCH	0.77	23.22	L15-βCH2		L15-δCH3'
	δCH3	0.43	22.01	L15-γCH		
	δCH3'	0.49	22.3	L15-γCH		L15-δCH3

Table 3.2. RKKD790_1592 fragment ions observed in the tandem mass spectra.

Fragment Ion	Fragment type	Observed [M+H]	Calculated Δppm
GFGSKPIDSFGLSWL	Parent	1592.8078	2.19
GFGSKPIDSFGLS	b13	1275.6355	1.02
GFGSKPIDSFGL	b12	1188.6041	0.59
GFGSKPIDSFG	b11	1075.5201	1.12
GFGSKPIDSF	b10	1018.4986	1.18
GFGSKPIDS	b9	871.4303	1.15
GFGSKPID	b8	784.3984	1.27
SFGLSWL	y7	809.4188	0.49
FGLSWL	y6	722.3868	0.55
GLSWL	y5	575.3183	0.87
LSWL	y4	518.2969	0.77
SWL	y3	405.213	0.49
WL	y2	318.1812	0.00
GFGSKPID+GLSWL	b8+y5	1358.7086	1.25
GFGSKPID+LSWL	b8+y4	1301.6867	1.61
GFGSKPID+SWL	b8+y3	1188.6041	0.59
GFGSKPID+WL	b8+y2	1101.572	0.73
GFGSKPIDS+GLSWL	b9+y5	1445.7408	1.04
GFGSKPIDS+LSWL	b9+y4	1388.7192	1.22
GFGSKPIDS+SWL	by+y3	1275.6355	1.02
GFGSKPIDS+WL	b9+y2	1188.6041	0.59

Marfey's Method was used to determine the stereochemistry of each amino acid[100]. Retention times of derivatized peptide hydrolysate were compared to that of authentic amino acid standards (Appendix, Figure 8.18, 8.19). The amino acid standards eluted at the following retention times: L-Lys: 12.37 min and 15.40, L-Ser: 14.34 min, L-Asp: 16.31 min, L-Pro: 22.08 min, L-Ile: 34.43 min, D-Ile: 40.58 min, L-Leu: 35.49 min, D-Leu: 41.17 min, L-Trp: 34.99 min and L-Phe: 35.57 min. Because of the presence of two free amine groups on L-Lys, FDAA can react with both creating two retention times for L-Lys. The FDAA derivatized peptide hydrolysate peaks were identified based on expected exact masses and eluted at 12.34 and 15.28, 14.24 min, 16.18 min, 22.00 min, 34.28 min, 35.38 min, and 35.57 min corresponding to L-Lys, L-Ser, L-Asp, L-Pro, L-Ile, L-Leu, and L-Phe respectively. Retention times corresponding to both D-Trp and L-Trp were present, so milder reaction conditions were employed to reduce racemization. Using a lower reaction temperature, tryptophan was derivatized without racemizing and a single diastereomer eluted at 35.35 min corresponding to L-Trp.

3.3.4 3D Molecular modelling of KG790_1592

To confirm that KG790_1592 was a member of the lasso peptide family, the 3D conformation was determined. To show that the linear tail was threaded through the ring, and the extent to which this occurs, computational modelling was used.

Six theoretical conformations were optimized and compared to the experimentally determined H-H distances. A total of 55 NOE parameters were compared and the total errors, the mean absolute deviation (MAD) and the maximum deviation are summarized in Table 3.3. Based on these results, the linear tail of KG790_1592 is likely threaded through the ring between Ser-13 and Trp-14 in a right hand conformation, which is

shown in Figure 3.3. These computational results agree with predictions based on mass spectrometry data where b8+y2 and b9+y2 fragments are observed. Interestingly, all other instances of lasso peptides previously reported exhibit the right handed conformation (loop type A) as well, which further supports the theoretical conclusions[84].

Table 3.3. Error values of H-H distances between experimentally determined interatomic distances and theoretical conformation predictions.

Candidate	Total Error (A)	MAD (A)	Max Deviation (A)	N > 1	N 1
ALeu	91.55	1.66	8.52	27	28
ASer	64.43	1.17	3.96	26	29
ATrp	63.82	1.16	4.07	22	33
BLeu	84.37	1.53	6.5	25	30
BSer	69.35	1.26	4.65	26	29
BTrp	78.74	1.43	7.18	25	30

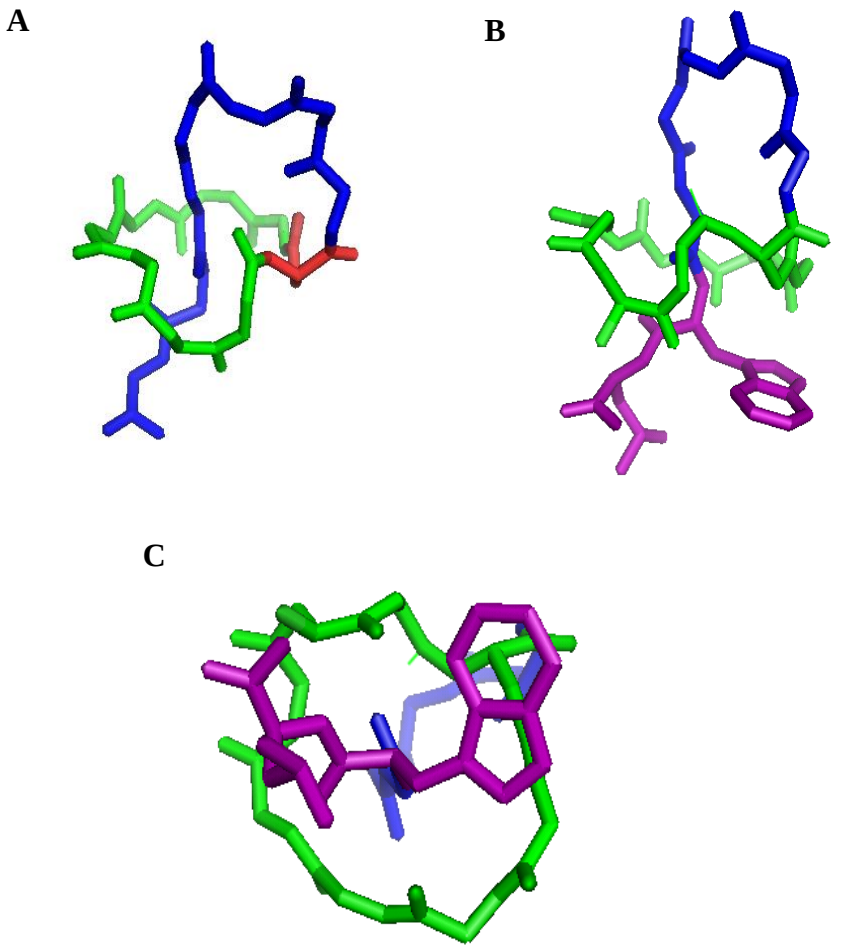


Figure 3.3. The peptide backbone of the 3D conformation of RKKD790_1592 determined by computational modelling.
The backbone of the ring portion is coloured green and the threaded linear tail in blue. **A**. Cyclization occurs between the amino group of Gly-1 and Asp-8 which is coloured red. **B**. Side view of the threaded conformation showing the bulky side chains of Trp-14 and Leu-15 in purple. **C**. Bottom view of the threaded structure showing the side chains of Trp-14 and Leu-15 in purple. The large side of the bulky groups relative to the size of the ring shows that the steric bulk is responsible for locking the linear tail in place.

The lack of cysteine residues in the structure of KG790_1592 causes it to be classified as a class II lasso peptide. Lasso peptides are a rapidly growing family of RiPPs. KG790_1592 consists of 15 amino acids and is the smallest lasso peptide that has been reported to date. It contains a ring consisting of 8 amino acids and a tail with 7 amino acids; 5 above the ring and 2 below. The bulky side chains of both Trp-14 and Leu-15 contribute to the steric lock responsible for holding the tail within the ring. Two amino acids is the shortest sequence located below the ring, and is only observed in KG790_1592 and microcin J25[90]. KG790_1592 also contains a loop consisting of 5 amino acids, which is shorter than the loop of many lasso peptides and is only one residue longer than the shortest observed loop regions in lariatin A[91] and RES-701-1[87]. The macrolactam ring of other class II lasso peptides is formed between glycine and either glutamic acid at position 8, or aspartic acid at position 9[85]. The ring in KG790_1592 is formed between glycine and aspartic acid at position 8, which is the first peptide with a confirmed lasso structure to contain this shorter linkage consisting of 25 atoms rather than 26 or 28. The combination of the small size of the ring, tail and loop region observed in the structure of KG790_1592 may provide an example of the minimum geometry required for a lasso fold. This example of the minimum structural features required to result in the stable and valuable lasso fold provides useful information for the characterization of this class of peptides and assists in the discovery of additional members.

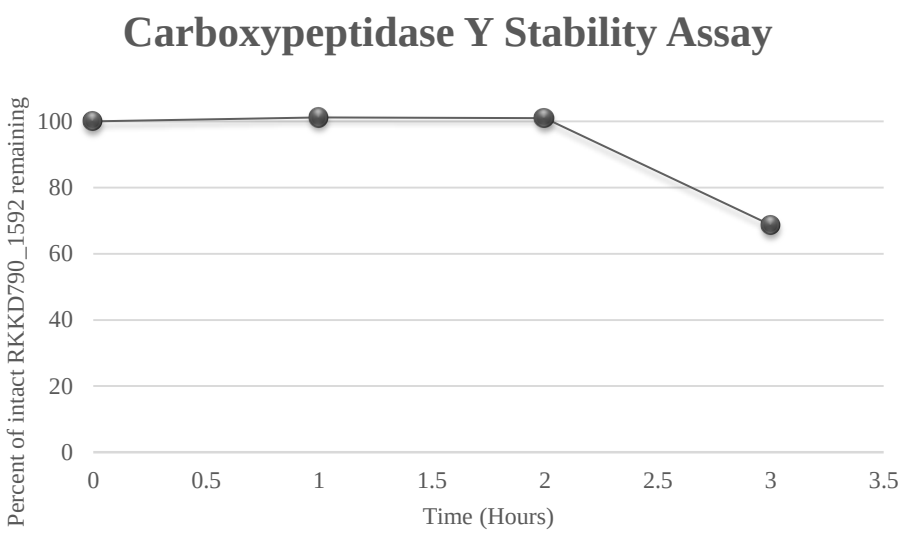
3.3.5 Stability assays

A significant challenge for becoming drug candidates that peptides face is a lack of stability in biological systems[97,98]. Lasso peptides are able to overcome this stability

issue because their unique threaded structure allows them to be more resistant to heat, chemical and enzymatic degradation than other peptides[85]. To evaluate the stability of RKKD790_1592 in physiological conditions, it was incubated with the peptide degrading enzyme, carboxypeptidase Y, and human serum. The amount of intact RKKD790_1592 after incubation was measuring using LC-HRMS at varying time points (Figure 3.4). When RKKD790_1592 was incubated with carboxypeptidase Y, approximately all of the RKKD790_1592 remained intact for two hours. After three hours of incubation RKKD790_1592 began to degrade and approximately 68 % of the original amount of RKKD790_1592 remained intact. Compared to linear peptides that degrade within minutes of incubation with carboxypeptidase Y, RKKD790_1592 was found to be extremely resistant to peptidase degradation.

RKKD790_1592 was also incubated with human serum to evaluate its physiological stability. The remaining amount of intact RKKD790_1592 was approximately 100 % for all time points that were evaluated, up to 48 hours. This shows the extreme stability of RKKD790_1592 in the presence of human serum, and shows that like other lasso peptides, the threaded structure of RKKD790_1592 causes it to be very resistant to degradation. This shows that RKKD790_1592 has to the potential to overcome the stability challenges of peptides associated to drug development challenges.

A



B

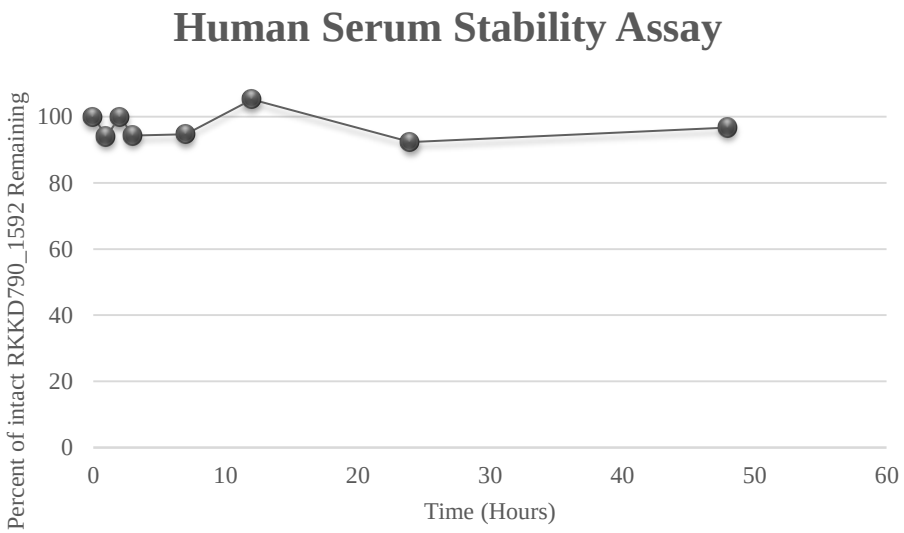


Figure 3.4. Evaluation of the physiological stability of RKKD790_1592 in the presence of A. Carboxypeptidase Y and B. human serum.

3.3.6 Biological activity and assays

KG790_1592 was evaluated for antimicrobial activity and cytotoxicity in multiple activity assays. No growth inhibition was observed at 50 µg/mL against methicillin-resistant *Staphylococcus aureus,* vancomycin-resistant *Enterococcus faecium, Staphylococcus warneri, Proteus vulgaris, Pseudomonas aeruginosa,* and *Candida albicans.*

Many lasso peptides have antibiotic activity against bacterial taxa that are closely related to the producing bacterium, and therefore KG790_1592 was evaluated for growth inhibition against *Arcanobacterium pyogenes, Corynebacterium pseudotuberculosis, Dermatophilus congolensis, Micrococus luteus, Rhodococcus equi* and *Mycobacterium smegmatis* in a disk diffusion assay. No zones of inhibition were observed in the presence of KG790_1592 for all 6 bacteria.

Cytotoxicity was also evaluated against a human breast adenocarcinoma and human foreskin BJ fibroblast cell lines. No growth inhibition was observed up to a concentration of 128 µg/mL for both the cancerous and normal cell lines.

Other class II lasso peptides that were found to have no antimicrobial activity include BI-32169[95], RES-701-1[87], caulosegnins I-III[93] and anantin[106]. Most of these lasso peptides that do not display antimicrobial activity possess very specific receptor antagonist or enzyme inhibition activity[85]. BI-32169 is a glucagon receptor antagonist[95], RES-701 is an antagonist of the endothelin type-B receptor[87] and anantin is an inhibitor of the atrial natriuretic factor[106]. KG790_1592 does not display any antimicrobial activity against several bacterial taxa, and no cytotoxicity against cancerous or healthy human cells. While the results of these assays do not reveal the potential applications of

KG790_1592, the lack of cytotoxicity shows the capability of having a very high therapeutic ratio. Biological activity may be highly specific and in combination with high stability and very low cytotoxicity, KG790_1592 may have the potential to be an effective therapeutic agent against a specific molecular target, or use as a moleulcar scaffold.

3.4 CONCLUSION

The discovery of KG790_1592 represented an exciting opportunity to describe a new member of the relatively new, but rapidly growing group of RiPPs, the lasso peptides. KG790_1592 possessed several structural features that made it unique compared to other members of this class of natural products. At the time of isolation it was the smallest lasso peptide that had been described, and allowed the boundaries for minimum size requirements to produce a lasso fold to be adapted to include KG790_1592.

KG790_1592 was evaluated for biological activity in several different bioassays. It displayed no antibiotic activity against any of the microbial pathogens that it was evaluated against, as well as no anticancer activity. However KG790_1592 was also not active against healthy human cell lines, and is extremely resistant to protease, heat and chemical degradation suggesting that it may have the potential for pharmaceutical applications if specific enzyme inhibition activity is found because of its stability and potential for a high therapeutic window.

The structure of this new lasso peptide was also discovered and described by Prof. Dong-Chan Oh's research group who isolated the same lasso peptide from a similar

Streptomyces sp. isolated from deep sea sediment off the coast of Jeju Island, Korea[107]. They named this lasso peptide sungsanpin and found that it can inhibit cell invasion of A549 lung cancer cells.

CHAPTER 4: ISOLATION, STRUCTURE ELUCIDATION OF CYSTARGAMIDE: A LIPOPEPTIDE NATURAL PRODUCT FROM *KITASATOSPORA CYSTARGINEA*

This chapter is a modification of the material published as

Gill, K.A., Berrue, F., Arens, J.C., and Kerr, R.G. Isolation and structure elucidation of cystargamide, a lipopeptide from *Kitasatospora cystarginea. J. Nat. Prod.* **2014,** *77(6),* 1372-1376.

4.1 INTRODUCTION

Peptides encompass an enormous class of natural products with great diversity in both structure and biological applications[108]. Peptide natural products are often split into two classes based on their biosynthetic origin. Peptides synthesized by the ribosome are classified as ribosomally synthesized post-translationally modified peptides (RiPPs)[83], and peptides that are biosynthesized by modular enzymes that are distinct from the ribosome are classified as non-ribosomal peptides (NRPs)[108]. NRPs are biosynthesized by nonribosomal peptide synthetases (NRPS), which are large multifunctional enzymes that have conserved modes of synthesis[108-110]. NRPS enzymes often incorporate non-proteinogenic amino acids and peptide modifications including methylation, hydroxylation, acylation, and glycosylation, which results in a vast amount of structural diversity observed in this class of natural products[111]. NRPs have also been observed as linear, cyclic and branched-cyclic, which further contributes to the structural diversity of these natural products[111,112].

An interesting and pharmaceutically significant group of acylated cyclic NRPs is the cyclic lipopeptides[112]. Several examples of members of this natural product group are shown in Figure 4.1. This group of NRPs is classified by the addition of a medium to long chain fatty acid that is attached to a cyclic or branched cyclic peptide. The fatty acid groups are often essential for biological activity. One of the most studied cyclic lipopeptides is daptomycin, which was isolated from *Streptomyces roseosporus*[113]and is approved by the FDA for treatment of skin and soft tissue infections and MRSA in the blood stream[38]. There are also several other lipopeptides with potent antibiotic activity including the Calcium Dependent Antibiotics (CDAs)[114], friulimicin[115],

laspartomycins[116], and surfactin[117]. Discovering new members of the cyclic lipopeptide group of natural products offers excellent opportunities to discover novel antibiotics that may be useful for combatting multi-drug resistant microbial pathogens.

Members of the genus *Kitasatospora* are classified as rare Actinobacteria and have been found to be prolific sources of bioactive secondary metabolites[46]. Natural products produced by this genus include the bafilomycins[118], talosins[119], and more recently the satosporins[76]. In this chaper, the purification, structure elucidation and biological activity of the novel peptide, cystargamide will be discussed. Cystargamide is a cyclic lipopeptide that is cyclized by an ester linkage between the C-terminus acid and the hydroxy group of a threonine side chain. Cystargolide contains a 2,3-epoxy fatty acid chain that is linked to the threonine amine. It has structural similarities to two lipopeptides, Q-6402 A and B, isolated from *Streptomyces* sp. Q-6402, which have been reported to inhibit phospholipase A2[120].

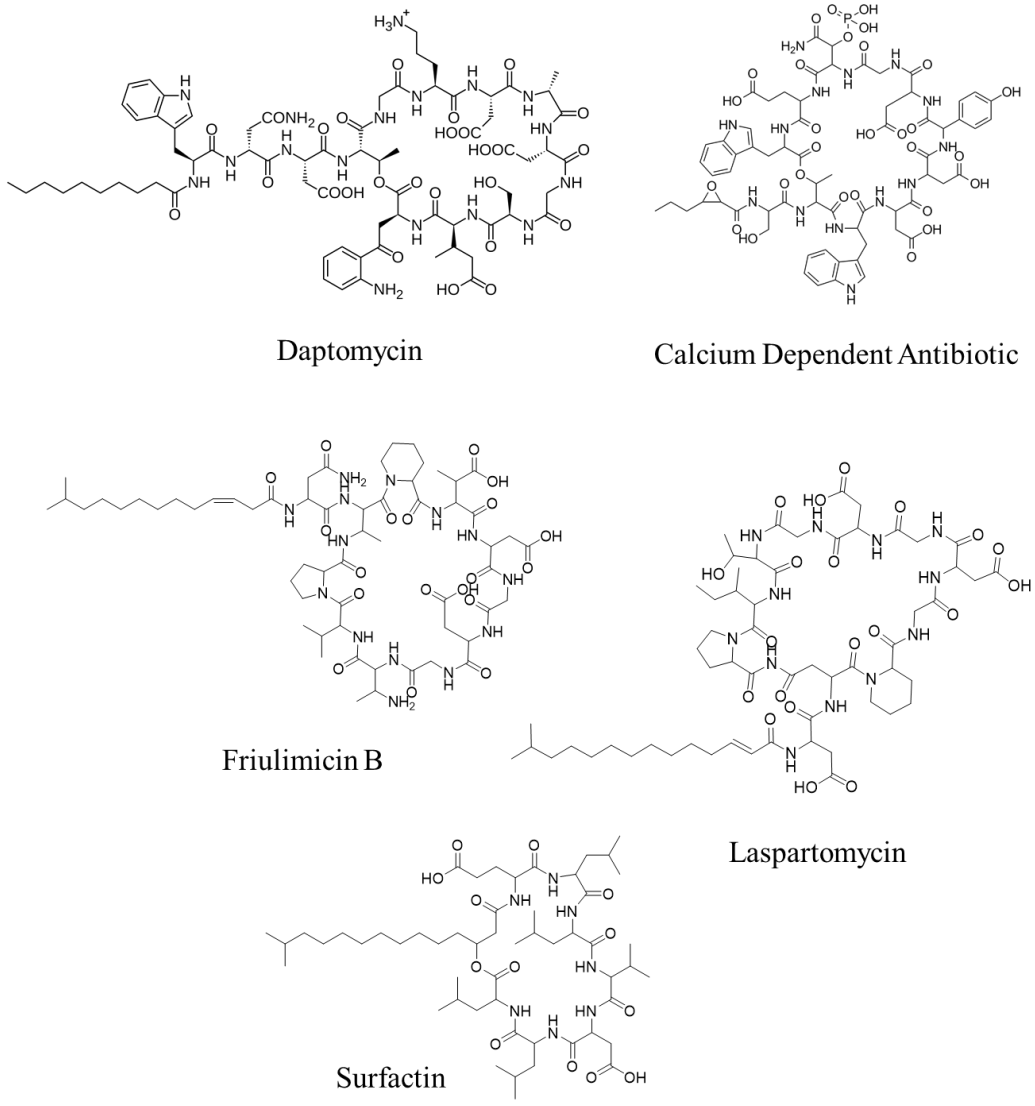


Figure 4.1. Structures of lipopeptide natural products with antimicrobial activity.

4.2 MATERIALS AND METHODS

4.2.1 Large scale fermentation, extraction and purification of cystargamide

From an ISP3[49] agar plate grown at 30ºC for 4-7 days, a single colony of *K. cystarginea* was used to inoculate 10 mL of seed medium (10 g/L glucose and 10 g/L yeast extract) containing 2-3 glass beads and fermented at 30 ˚C with agitation of 200 rpm for 48 hours. After 48 hours, 50 mL of seed medium in a 250 mL flask containing 3-5 glass beads was inoculated with 2.5 mL of the first seed culture and incubated under the same conditions for 48 hours. Five, 2.8 L fernbach flasks containing 1 L each of lean production medium (0.4 g/L glucose, 0.8 g/L galactose, 0.8 g/L maltose, 1.6 g/L dextrin, 0.8 g/L Bacto-soytone and 0.3 g/L ammonium sulfate) and 5-10 glass beads were inoculated with 50 mL each of second seed cultures and were incubated at 30˚C with agitation of 200 rpm for 72 hours. Each fermentation flask was extracted with 500 mL and then 300 mL of EtOAc, which were then partitioned twice with equal volume of diH_2O. After evaporation, the combined extract was dissolved in 10 mL of 80 % aqueous acetonitrile and partitioned with an equal volume of hexanes. The aqueous acetonitrile phase was evaporated to dryness *in vacuo.*

The dried acetonitrile fraction was fractionated using automated reversed-phase flash chromatography with a 15.5 g C_{18} column (High Performance GOLD, RediSep Rf) and eluted at 30 mL/min with a linear gradient from 50 % H_2O (solvent A1) and 50 % MeOH (solvent B1) to 100 % solvent B1 over 15 min. Cystargamide eluted at 10–11.5 min. Cystargamide was purified using RP-HPLC with a Luna 110 Å phenyl hexyl column (5 μm, 250 mm × 10 mm, Phenomenex). An isocratic elution with 30 % H_2O/0.1

% formic acid (solvent A2) and 70 % MeOH/0.1 % formic acid (solvent B2) over 20 min was used to elute cystargamide at 13.5 min.

4.2.2. Structure elucidation of cystargamide

Cystargamide: Yellow powder; $[\alpha]^{25}_{D}$ +7.4 (*c* 0.22, MeOH); ^{1}H and ^{13}C NMR data see Table 4.1. (+) HRESIMS *m*/*z* 954.4287 $[M + H]^+$, (calcd for $C_{49}H_{60}N_7O_{13}$, 954.4244).

Optical rotations were measured on a Rudolph Autopol III polarimeter using a 50 mm microcell (1 mL). The infrared spectrum was recorded using attenuated total reflectance, on a Bruker Alpha FT-IR spectrometer. NMR spectra were obtained on a 600 MHz Bruker Avance III NMR spectrometer. Chemical shifts (δ) were referenced to the DMSO-d_6 residual peaks at δ_H 2.50 ppm and δ_C 39.51 ppm. LC–HRMS was performed using a Thermo Q Exactive HPLC instrument with hyphenated MS-ELSD-UV detection: Finnigan LXQ ion trap mass spectrometer fitted with an ESI source, PDA, and LT-ELSD Sedex 80. Tandem MS spectra were measured on a Thermo LTQ Orbitrap Velos mass spectrometer. Automated flash chromatography was performed on a Teledyne Combiflash Rf200 using C_{18} RediSep columns. HPLC purifications were carried out on a Thermo Surveyor coupled with an evaporative light-scattering detector Sedex 55.

4.2.3 Stereochemical analysis by Marfey's Method

Amino acid absolute configurations were determined by Marfey's method of hydrolyzed cystargamide[100]. To a microconical vial, 200 µg of cystargamide and 250 µL of HCl (6M) was added and heated to 70-80 ºC for 90 minutes. 1 mL of 1N $NaHCO_3$ followed by 20 µL of 1-fluoro-2,4-dinitrophenyl-L-alanine (L-FDAA) (10 mg/mL in acetone) were added to the reaction vial and heated at 30-40 ºC for one hour before quenching with 100 µL of 6M HCl. The reaction mixture was reduced in volume under

air, then diluted to 1 mL with 50:50 MeOH/H_2O for LC-MS analysis. The resulting reaction mixture was analyzed by LC-MS using an LTQ Orbitrap Velos (ThermoScientific) HPLC system with a Hypersil Gold 100 Å column (Thermo, 1.9 μm C_{18} 50 mm × 2.1 mm). A linear gradient of 95 % H_2O/0.1 % formic acid (solvent A3) and 5 % CH_3CN/0.1 % formic acid (solvent B3) to 40 % solvent B3 over 55 min followed by a rapid increase to 100 % solvent B3 over 2 min then a hold for 3 min was used, with a flow rate of 400 μL/min. Retention times were compared with those of authentic derivatized standards to determine the amino acid configurations.

Hydroxytryptophan degraded during acid hydrolysis involved in Marfey's Method so an ozonolysis reaction was used to convert the hydroxyindole group to a more stable carboxylic acid so that Marfey's Method could be used. To a reaction vial, 700 μg of cystargamide was added with 5 mL of MeOH. The solution was cooled to −78 °C, and ozone in oxygen was bubbled through the solution for 30 min, followed by addition of 20 drops of 38 % H_2O_2 to quench the reaction. The reaction was heated to room temperature and evaporated under N_2 until dry. Acid hydrolysis and derivatization with L-FDAA were repeated on the ozonolysis product followed by LC-MS analysis as described above.

4.2.4 Antimicrobial assays

Cystargamide was tested for antimicrobial activity against the human microbial pathogens, methicillin-resistant *Staphylococcus. aureus* ATCC 33591 (MRSA), vancomycin-resistant *Enterococcus. faecium* EF379 (VRE), *Staphylococcus warneri* ATCC 17917, and *Candida. albicans* ATCC 14035. All assays were carried out according to Clinical Laboratory Standards Institute testing standards in a 96-well plate

and in triplicate[53]. Control antibiotics used were vancomycin for MRSA and *S. warneri*, rifampicin for VRE, or nystatin for *C. albicans*. Cell growth was determined by measuring the optical density at 600 nm (OD_{600}) at time zero and after incubation for 22 hours at 37 ºC using a Thermo Scientific Varioskan Flash plate reader. The change in OD_{600} after incubation was compared to untreated negative control wells and plotted as percent growth inhibition. Cystargamide was evaluated at six concentrations ranging from 2 µg/mL to 64 µg/mL.

Cystargamide was tested for antibiotic activity against *Bacillus subtilis* NRRL-B-558 in a disk diffusion assay. Assay conditions were performed at cystargamide concentrations of 2 µg/disk, in the presence and absence of 12 mM Ca^{2+}. Penicillin G was used as a positive control. Zones of inhibition were measured with calipers after 24 h of growth.

4.2.5 Phospholipase A2 inhibition assay

Cystargamide was also evaluated for inhibition against human secretory phospholipase A2 type V. The assay was performed using the sPLA2 (Type V) inhibitor screening assay kit (Cayman Chemical Item No. 10004883) and the recommended protocol. The assay kit included sPLA2 assay buffer 10X (Item No. 765010), sPLA2 DTNB (Item No. 765012), sPLA2 diheptanoyl thio-PC substrate (Item No. 765014), sPLA2 human type V (Item No. 10004913), 96-well solid plate for colorimetric assay and cover sheet (Item No. 400014, 400012). The assay was performed at room temperature, in triplicate and absorbance at 415 nm was measured using a Spectra Max M2 (Molecular devices) plate reader. Absorbance of cystargamide treated wells were compared to the untreated negative control wells and used to calculate the percent

enzyme activity at nine concentrations of cystargamide ranging from 0.16 µM to 40 µM. Thioetheramide-PC was used as a positive control at four concentrations ranging from 0.62 µM to 40 µM.

4.2.6 Activation of microbial secondary metabolites

To evaluate the effects of cystargamide on the growth and metabolite production of other actinomycetes, a disk diffusion assay was used. To a paper disk, 50 µL of cystargamide (10 mg/mL in EtOH) was added and allowed to dry. Dry disks were applied to ISP2[49] agar plates that were inoculated with 100 µL of a *Streptomyces coelicolor* spore suspension ($1x10^6$ spores/mL). The same concentration of surfactin was used as a positive control, and EtOH as a negative control. Plates were incubated at 30 ºC for 7 days, and all conditions were repeated in triplicate.

4.3 RESULTS AND DISCUSSION

4.3.1 Fermentation, extraction and isolation of cystargamide

Through the use of LC-MS/PCA metabolomics screening method described in Chapter 2, it was identified that *Kitasatospora cystarginea* produced three putatively novel metabolites. This led to the selection of *K. cystarginea* for fermentation scale up. A 5 L scale-up fermentation was undertaken and culture broths were extracted with ethyl acetate. The resulting organic extract was fractionated by C_{18} liquid chromatography and further purified using RP-HPLC with a phenyl hexyl stationary phase to yield pure cystargamide. Several additional large scale fermentations were required to purify enough material for all structural characterizations and biological assessments.

4.3.2 Structure elucidation of cystargamide

Cystargamide was isolated as a yellow solid, and HRMS supported the molecular formula $C_{49}H_{59}N_7O_{13}$ indicating 24 degrees of unsaturation. NMR data (Table 4.1) revealed the presence of six amide protons which was indicative of a peptidic structure. Seven distinct ^{1}H spin systems were identified as six amino acid side chains and one oxidized fatty acid chain by interpretation of the TOCSY spectrum (Figure 4.2). Amino acid side chains were assigned primarily using ^{1}H, HSQC, COSY, and HMBC spectra, and the amino acid sequence was assigned using the NOESY and HMBC spectra (Appendix, Figures 8.20-8.30).

The glycine residue was assigned on the basis of the COSY correlation between Gly-NH (δ_H8.22)/H$_2\alpha$ (δ_H 4.32, 3.56). The 4′-hydroxyphenylglycine residue (Hpg) was assigned on the basis of COSY correlations between Hpg-NH (δ_H 8.72)/Hα (δ_H 5.21); H-2′ (δ_H 6.84)/H-3′ (δ_H 6.60); and the HMBC correlations from Hα to C-1′ (δ_C 127.7), C-2′ (δ_C 129.1); H-2′ to C-1′; and H-3′ to C-4′ (δ_C 156.3). A 4′-hydroxyphenolic group was supported by the characteristic ^{13}C chemical shift at δ_C 156.3 and both ^{1}H signals (H-2′ and H-3′) that each integrated for 2H. The 5-hydroxytryptophan (Htrp) was identified on the basis of COSY correlations between Htrp-NH (δ_H8.74)/Hα (δ_H 4.33); Hα/H$_2\beta$ (δ_H 3.00, 2.93); Htrp H-1 (δ_H 10.52)/H-2 (δ_H 7.04); H-7 (δ_H 7.08)/H-6(δ_H 6.57) in addition to HMBC correlations from H$_2\beta$ to C-3a (δ_C 127.6), C-3 (δ_C 107.9), C-2 (δ_C 123.9); Htrp H-1 to C-7a (δ_C 130.0), C-3; H-2 to C-7a, C-3a, C-3; H-7 to C-5 (δ_C 149.9), C-3a; H-6 to C-7a, C-4; H-4 (δ_H 6.85) to C-6, C-3, C-7a. The ^{13}C resonance at δ_C 149.9 (C-5) was in agreement with a hydroxy substituent at the C-5 position. The glutamic acid residue was assigned by the interpretation of the TOCSY spectrum that

revealed the spin system Glu-NH (δ_H7.09)/Hα (δ_H 4.31)/H_2β (δ_H 2.04, 1.92)/H_2γ (δ_H 2.36, 2.29) in addition to the key HMBC correlation from H_2γ to COOH (δ_C 175.4). The phenylalanine residue was determined on the basis of COSY correlations that indicated the two spin systems: Phe-NH (δ_H 8.75)/Hα (δ_H4.47)/H_2β (δ_H 3.15, 2.72) and H-2′ (δ_H 7.16)/H-3′ (δ_H 7.19)/H-4′ (δ_H 7.13); and the HMBC correlations from H_2β to C-1′ (δ_C 137.7), C-2′ (δ_C 128.8); H-2′ to C-4′ (δ_C 126.0), and H-3′ to C-1′. The key COSY correlations between Thr-NH (δ_H 7.68)/Hα (δ_H 4.55), Hα/Hβ (δ_H 5.40), and Hβ/H_3γ (δ_H 1.10) confirmed the presence of the amino acid, threonine. Finally, the interpretation of the TOCSY data revealed the last spin system H-2 (δ_H 3.42)/H-3 (δ_H 2.8)/H_2-4 (δ_H 1.56, 1.50)/H_2-5 (δ_H 1.37)/four overlapping methylenes (δ_H 1.30–1.26)/H_3-10 (δ_H 0.89). The downfield resonances for H-2 and H-3 were characteristic of methine signals bearing a heteroatom. The COSY correlation between H-2 and H-3, the HMBC correlation from H-2 to Epd-CO and the NOESY correlation between Thr-NH and H-2 indicated the presence of a 2,3-epoxide ring. The HRMS data supported this structure assignment on the basis of the additional degree of unsaturation and mass of an oxygen atom that were not yet attributed to the structure.

Table 4.1. ^{1}H (600 MHz) and ^{13}C (150 MHz) NMR Data for cystargamide in DMSO-d_6.

Position		δ_C, type	δ_H (J, Hz)	COSY	NOESY	HMBC (H→C)
Epd[a]	CO	167.5, C				
	2	53.3, CH	3.42, d (1.8)	Epd-3	Thr-NH, Epd-4	Epd-CO
	3	57.5, CH	2.80, ddd (5.5, 5.5 1.8)	Epd-2,Epd-4	Thr-NH, Epd-4	
	4	30.5, CH_2	1.56, m; 1.50, m	Epd-3, Epd-5		
	5	25.0, CH_2	1.37, m	Epd-4, Epd-6		
	6[d]	28.5, CH_2	1.30, m			
	7[d]	21.9, CH_2	1.28, m			
	8[d]	28.7, CH_2	1.26, m			
	9[d]	31.1, CH_2	1.26, m			
	10	13.7, CH_3	0.89, t (6.9)	Epd-9		
Thr	CO	168.0, C				
	NH		7.68, d (8.9)	Thr-α	Epd-2, Epd-3	Epd-CO
	α	54.1, CH	4.55, dd (9.0, 1.5)	Thr-NH, Thr-β	Phe-NH	Thr-CO, Epd-CO
	β	70.2, CH	5.40, m	Thr-α, Thr-γ	Phe-NH	Gly-CO
	γ	16.2, CH_3	1.10, d (6.3)	Thr-β		Thr-α, Thr-β
Phe	CO	169.8, C				
	NH		8.75, m	Phe-α	Thr-α, Thr-β	
	α	54.2, CH	4.47, ddd (9.3, 3.7, 3.7)	Phe-NH,Phe-β	Glu-NH	Phe-CO, Thr-CO
	β	36.8, CH_2	3.15, dd (13.9, 3.6); 2.72, dd (13.8, 3.5)	Phe-α	Phe-2’	Phe-1’, Phe-2’, Phe-CO
	1’	137.7, C				
	2’	128.8, CH	7.16, m	Phe-3’	Phe-NH, Phe-α, Phe-β	Phe-4’, Phe-β
	3’	127.8, CH	7.19, m	Phe-2’, Phe-4’		Phe-1’
	4’	126.0, CH	7.13, m	Phe-3’		Phe-2’
Glu	CO	168.1, C				
	NH		7.09, m	Glu-α	Phe-α	
	α	51.6, CH	4.31, m	Glu-NH, Glu-β		Glu-CO
	β	28.3, CH_2	2.04, m 1.92, m	Glu-α, Glu-γ	Htrp-NH	
	γ	31.8, CH_2	2.36, m; 2.29, m	Glu-β		Glu-COOH
	COOH	175.4, C				

Position		δ_C, type	δ_H (J, Hz)	COSY	NOESY	HMBC (H→C)
	CO	171.2, C				
	NH		8.74, m	Htrp- α	Htrp-2, Glu-α, Glu-β	
	α	55.1, CH	4.33, m	Htrp-NH, Htrp-β	Hpg-NH	Htrp-CO, Glu-CO
	β	26.2, CH_2	3.00, dd (14.1, 7.8); 2.93, dd (14.2, 6.8)	Htrp- α	Htrp-4, Htrp-2	Htrp-CO, Htrp-3a, Htrp-3, Htrp-2
	1		10.52, s	Htrp-2		Htrp-3a, Htrp-7a, Htrp-3, Htrp-2
Htrp[b]	2	123.9, CH	7.04, s	Htrp-1	Htrp-1	Htrp-3a, Htrp-7a, Htrp-3
	3	107.9, C				
	3a	127.6, C				
	4	102.2, CH	6.85, d (2.3)		Htrp-β, Htrp-α	Htrp-3, Htrp-5, Htrp-6, Htrp-7a
	5	149.9, C				
	6	111.0, CH	6.57, dd (8.6, 2.2)	Htrp-7		Htrp-7a, Htrp-4
	7	111.3, CH	7.08, d (8.6)	Htrp-6		Htrp-5, Htrp-3a
	7a	130.0, C				
	CO	169.9, C				
	NH		8.72, m	Hpg-α	Gly-NH, Hpg-2', Htrp-α	
	α	55.8, CH	5.21, d (8.4)	Hpg-NH	Gly-NH, Hpg-2',	Htrp-CO, Hpg-CO, Hpg-1', Hpg-2'
Hpg[c]	1'	127.7, C				
	2'	129.1, CH	6.84, d (8.9)	Hpg-3	Hpg-NH, Gly-NH, Hpg-α	Hpg-4', Hpg-α
	3'	114.6, CH	6.60, d (8.7)	Hpg-2		Hpg-4', Hpg-1'
	4'	156.3, C				
	CO	168.2, C				
	NH		8.22, d (5.8)	Gly-α	Hpg-NH, Hpg-α, Hpg-2'	Hpg-CO
Gly	α	40.0, CH_2	4.32, m; 3.56, m	Gly-NH		Gly-CO, Hpg-CO

[a]Epd: 2,3-epoxydecanoyl, [b]Htrp 5'-hydroxytryptophan, [c]Hpg: 4'-hydroxyphenylglycine, [d]overlapping signals may be interchanged.

The NOESY signals between Gly-NH/Hpg-Hα; Hpg-NH/Htrp-Hα; Htrp-NH/Glu-Hα; Glu-NH/Phe-Hα; Phe-NH/Thr-Hα revealed the sequence of amino acids to be Gly-Hpg-Htrp-Glu-Phe-Thr. The HMBC correlation from Thr-Hβ to Gly-CO (δ_C 168.2) as well as the chemical shift of Thr-Hβ suggested that Gly-Thr were connected in a cyclic structure via an ester linkage. The epoxydecanoyl group was linked to the peptidic moiety according to the strong NOESY correlation between Thr-NH/H-2, as well as the key HMBC correlation from Thr-NH, Thr-Hα to Epd-CO (δ_C 167.5). The molecular structure of cystargamide and key NMR correlations used for structure elucidation are shown in Figure 4.2. The planar structure of cystargamide, and in particular the amino acid sequence, was confirmed by tandem mass spectrometry.

Cyclic lipopeptides are characterized by the linkage of a polypeptide moiety to a fatty acid side chain. They often vary greatly in size, amino acid, and fatty acid composition and commonly contain non-proteinogenic amino acids[112]. Incorporation of a large assortment of fatty acid building blocks in combination with a number of non-proteinogenic amino acids results in vast chemical diversity for this group of natural products. The 4′-hydroxyphenylglycine (Hpg) residue in cystargamide, is found in a number of different classes of natural products, including lipopeptides such as the CDAs[114], enduracidins[121], and ramoplanins[122]. However, the amino acid residue 5-hydroxytryptophan (Htrp) is not as commonly incorporated, especially in microbial natural products. Htrp is a naturally occurring mammalian metabolite that is a biosynthetic precursor to the endogenous neurotransmitters serotonin and melatonin, and is rarely found in microbial natural products[123].

Like other members of the lipopeptides including daptomycin and the CDAs, cystargamide is cyclized by an ester linkage between the C-terminal acid and the hydroxy group of a threonine side chain[112]. Cystargamide also contains a rare 2,3-epoxy fatty acid side chain that has only been reported once in the CDA type-lipopeptides CDA1, 2, 3b, and 4b[114]. It is worth noting that these CDAs possess the same *trans* configuration for the epoxide ring, but in this case, in a shorter fatty acid chain[124]. Therefore to the best of our knowledge, cystargamide is the first report of a natural product containing a 2,3-epoxydecanoyl fatty acid side chain.

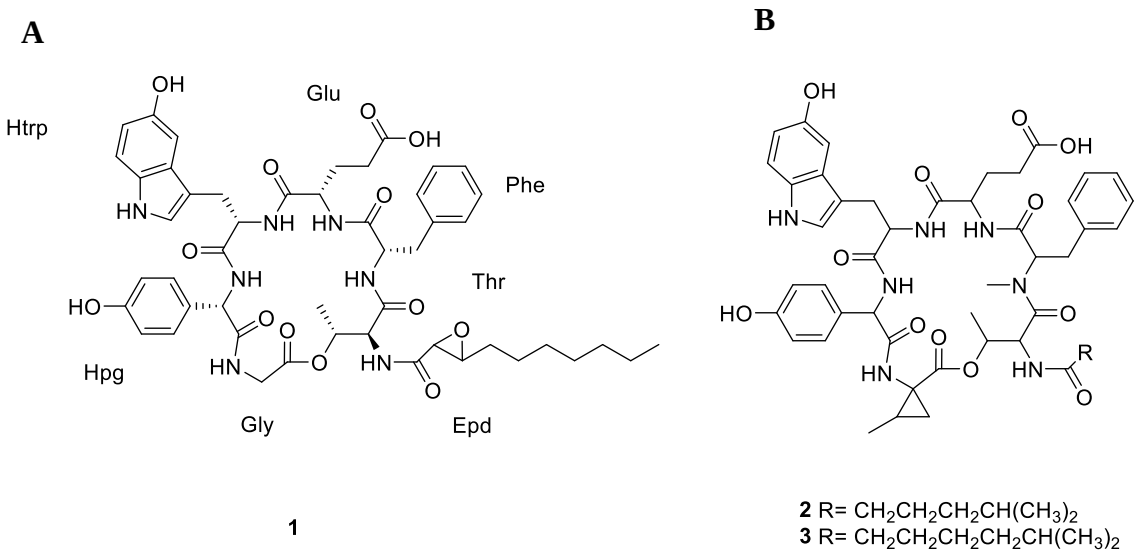
4.3.3 Stereochemical analysis by Marfey's Method

The absolute configurations of the amino acids were determined using Marfey's method[100]. Cystargamide was hydrolyzed using HCl and then derivatized with *N*-(2,4-dinitro-5-fluorophenyl)-L-alaninamide (L-FDAA) and analyzed by LC-MS; retention times were then compared with those of derivatized standards (Appendix, Figures 8.23-8.24). Standard amino acids eluted at 35.20 min (L-Phe), 18.00 min (L-Glu), 13.83 min (L-Hpg), 14.97 min (L-Thr), 20.60 min (D-Thr) and 15.50 and 17.67 min (DL-*allo*-Thr). Hydrolyzed cystargamide eluted at 35.25 min (L-Phe), 18.00 min (L-Glu), 13.82 min (L-Hpg) and 14.94 min (L-Thr).

All derivatized amino acids, with the exception of Htrp, eluted at retention times similar to those of their corresponding L-standards, revealing that cystargamide contained L-Hpg, L-Glu, L-Phe, and L-Thr. The Htrp residue was not detected, likely due to its lability; therefore cystargamide was reacted with ozone using an oxidative workup, followed by acid hydrolysis, derivatization with L-FDAA and analysis by LC-MS. Standard derivatized amino acids eluted at 16.28 min (L-Asp) and 18.73 min (D-Asp).

The derivatized ozonolysis product eluted at 16.30 min revealing L-Asp, which corresponded to the presence of L-Htrp in cystargamide.

The relative configuration of the epoxide ring was determined by using the J_{HH} coupling constant between H-2 and H-3. A small value of 1.8 Hz indicated that the vicinal protons had a *trans* relationship, revealing the relative configuration to be *2*R*, *3*S*[125]. This was also supported by a strong NOESY correlation between H-2 and H-4.



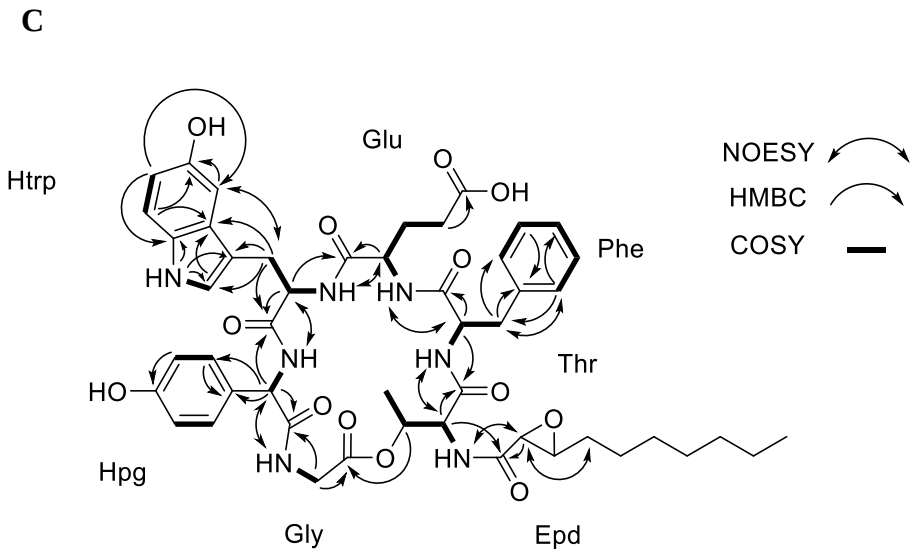


Figure 4.2. **A.** Molecular structure of cystargamide (**1**).
B. The molecular structure of Q-6402 a (**2**) and b (**3**)[120]. **C.** Key COSY, NOESY and HMBC correlation used to elucidate the structure of cystargamide.

4.3.4 Antimicrobial activity

Due to the structural similarity between cystargamide and antibiotic lipopeptides like CDA, the antimicrobial activity of cystargamide was evaluated against methicillin-resistant *Staphyloccocus aureus* (MRSA), vancomycin-resistant *Enterococcus faecium* (VRE), *Staphylococcus warneri*, and *Candida albicans*. Cystargamide displayed no significant growth inhibition against all four pathogens up to a concentration of 64 μg/mL, which is displayed in Figure 4.3. Cystargamide was also tested in a disk-diffusion assay with *Bacillus subtilis* in the presence and absence of calcium to determine if calcium was required for activity, like what is observed for CDA. Cystargamide displayed no antibiotic activity under either condition up to 2.0 μg/disk. Although CDA and cystargamide are both cyclic lipopeptides that contain a similar 2,3-epoxy fatty acid, CDA contains a total of eleven amino acids with ten comprising the cyclic portion, while cystargamide only consists of six cyclized amino acids. They also have several differences in amino acid position around the macrocycle. These structural differences between CDA and cystargamide are likely the reason that cystargamide did not display similar antimicrobial activity compared to CDA.

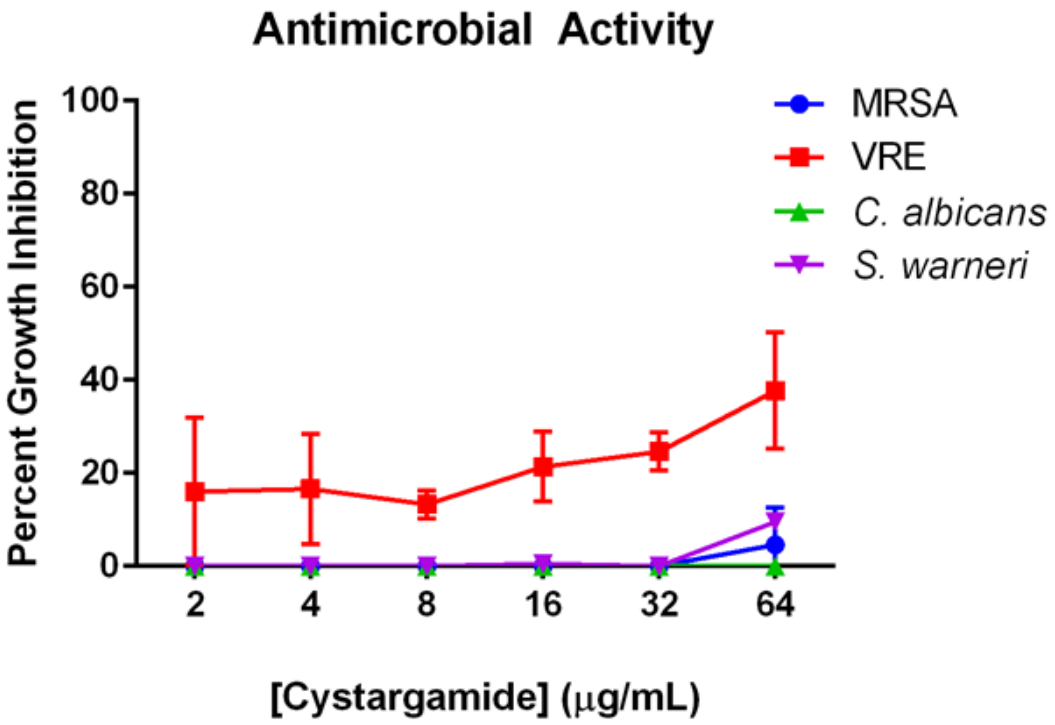


Figure 4.3. The percent growth inhibition of cystargamide against four human microbial pathogens.

4.3.5 Phospholipase A2 inhibition assay

Cystargamide shares significant structural similarities with the lipopeptides Q-6402 A and B, however these compounds have only been reported in a conference abstract and there is no detailed spectroscopic information available[120]. Cystargamide and Q-6402 A and B differ in the structure of the fatty acid chain, by the presence of the non-proteinogenic residue 2-methyl-1-amino-cyclopropanecarbocylic acid (MeACC) instead of glycine, and N-methylation of the phenylalanine residue (Figure 4.2). Q-4602 A and B were reported as inhibitors of rabbit platelet PLA2 with IC_{50} values of 9 µM and 6 µM for A and B respectively. Q-6402 B also exhibited anti-inflammatory activity with an ED_{50} value of 121 µM/ear in a mouse ear adema assay. When cystargamide was evaluated in a PLA2 inhibition assay it showed very weak activity, with an IC_{50} value greater than 40 µM (Figure 4.4). The difference in potency of cystargamide compared to Q-6402 A and B could be due to the structural differences, but it is more likely a result of different isoforms of the enzyme PLA2 that were used. Rabbit platelet PLA2 was used to evaluate inhibition of Q-4602 A and B, while human secretory type V PLA2 was used for cystargamide. Human secretory type V PLA2 was selected for the assay despite the differences to rabbit platelet PLA2 because the human isozyme is a more clinically relevant target to evaluate anti-inflammatory activity, and would result in a better indication in the potential *in vivo* anti-inflammatory activity of cystargamide. The lack of inhibition of gainst human PLA2 suggests that cystargamide is not a candidate for development as an anti-inflammatory agent.

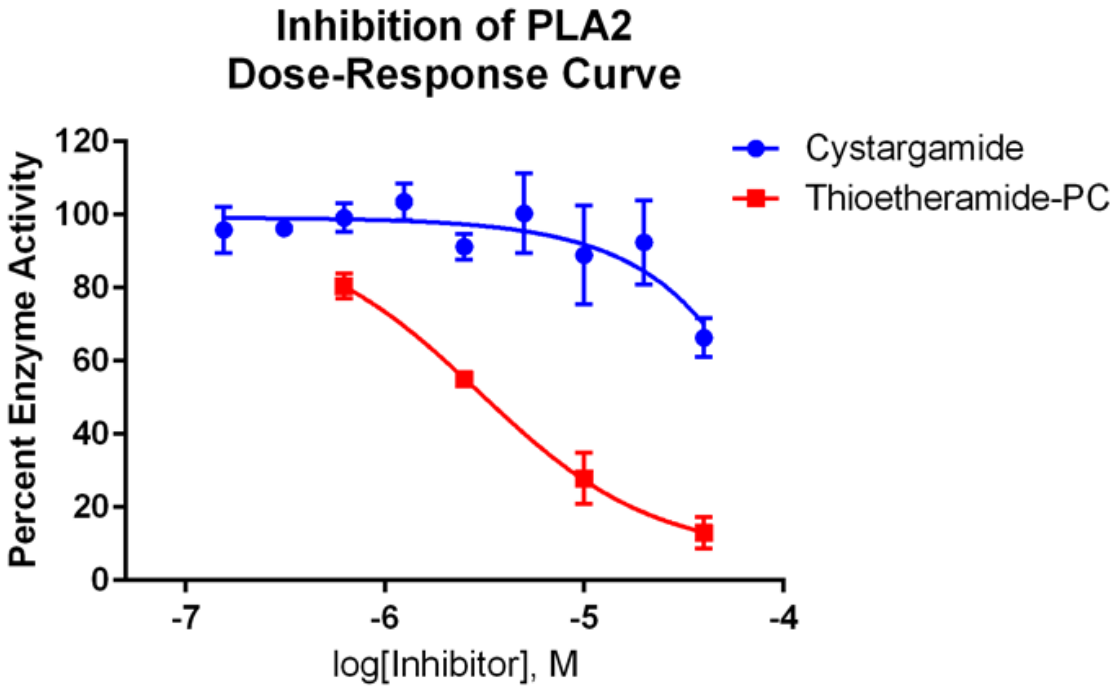


Figure 4.4. A dose-response curve showing the percent enzyme inhibition of cystargamide against human sPLA2 type V.

4.3.6 Activation of microbial secondary metabolites

The lipopeptide surfactin has been shown to inhibit development of aerial mycelia and spores in *Streptomyces coelicolor*[126]. Since cystargamide is also a lipopeptide, a disk diffusion assay was used to determine whether cystargamide has similar effects on the growth cycle of *S. coelicolor*. When *S. coelicolor* was grown in the presence of cystargamide, it did not inhibit the growth of aerial mycelia, however a clear difference in the metabolite production compared to the negative control was observed. On surfactin and EtOH control plates, no pigments were produced, however on plates containing cystargamide, a dark purple pigment was observed in all replicates, which can be seen in Figure 4.5. The identity of these pigments were confirmed as actinorhodin and undecylprodigiosin using the LC-MS "chemical screening" method described previously (Chapter 2)[127]. This shows that cystargamide acted as an inducer of secondary metabolite production in the model organism *S. coelicolor* and may play a role in microbial communication as a signalling molecule. These preliminary results show that the addition of cystargamide to microbial fermentations for natural products discovery programs may be a useful strategy to activate secondary metabolite biosynthesis and access cryptic metabolites. This is very useful for the discovery of novel natural products that are otherwise inaccessible under standard laboratory techniques.

A

B

C

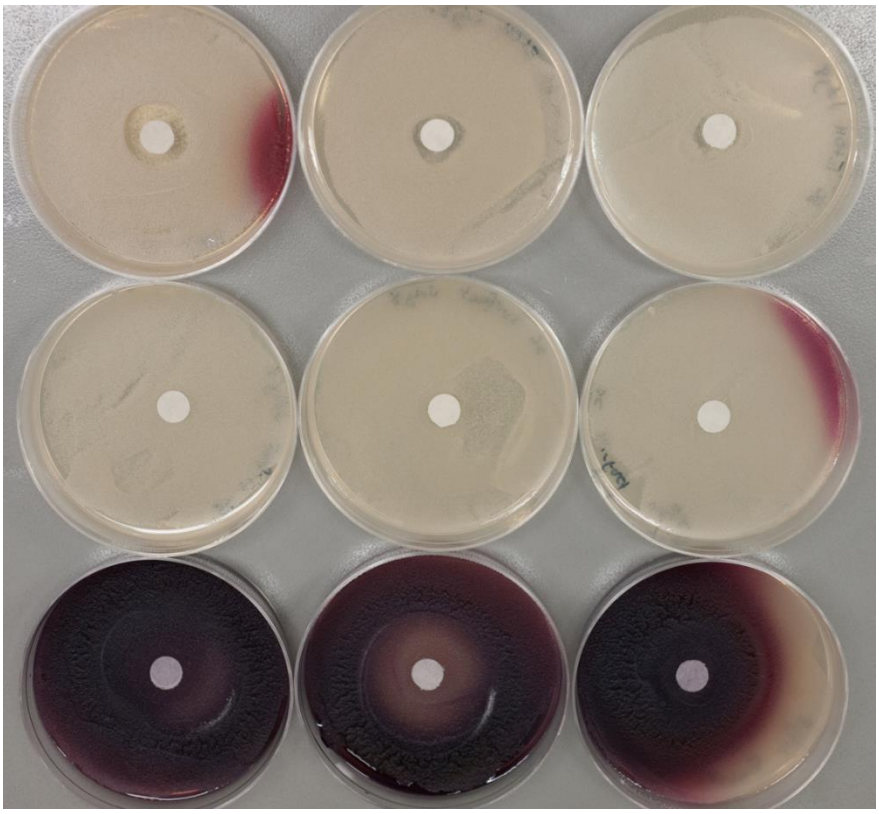


Figure 4.5. A disk diffusion assay showing *Streptomyces coelicolor* grown in the presence of **A.** ethanol, **B.** surfactin and **C.** cystargamide.

4.4 CONCLUSION

A novel lipopeptide natural product was isolated from the fermentation broth of a rare actinomycete *K. cystarginea.* The discovery of new natural products from this genus of bacteria shows that they are an excellent source for the isolation of novel bioactive metabolites. To the best of our knowledge, this is the first report of a natural product that contains a 2,3-epoxydecanoyl fatty acid coupled to a peptide. Cystargamide also contains the non-proteinogenic amino acids 4-hydroxyphenylglycine and the rare 5-hydroxytryptophan. Lipopeptide natural products are a very complex and diverse family of natural products and the addition of cystargamide to this group adds unique features to the structural possibilities that bacteria are capable of biosynthesizing.

Cystargamide was evaluated for biological activity in several bioassays. While it showed little or no antimicrobial activity or inhibition of human sPLA2 enzyme, it showed promising activity in a preliminary secondary metabolite induction assay with the model organism *S. coelicolor.* The use of cystargamide as an inducer of natural product biosynthesis in actinomycetes is an exciting opportunity to access cryptic natural products that may have useful applications to benefit human health.

CHAPTER 5: ISOLATION STRUCTURE ELUCIDATION AND BIOLOGICAL ACTIVITY OF CYSTARGOLIDES A AND B: NOVEL β-LACTONE CONTAINING PEPTIDES FROM *KITASATOSPORA CYSTARGINEA*

This chapter is a modification of the material published as

Gill, K.A. Berrue, F., Arens, J.C., Carr, G. and Kerr, R.G. Cystargolides, 20S proteasome inhibitors isolated from *Kitasatospora cystarginea. J. Nat. Prod.* **2015**, *78(4)*, 822-826.

5.1 INTRODUCTION

Natural products have been an excellent resource to explore for the discovery of pharmaceutical agents, especially with antibiotic and anticancer properties[14]. From the 1940s to 2014, 75 % of all FDA approved anticancer drugs were derived in some way from natural products and are classified as members of groups other than synthetic including natural products, natural product mimics and derivatives, botanicals, biologics and synthetics with natural pharmacophores[11]. Unaltered natural products and their derivatives account for 49 % of all FDA approved anticancer drugs. This remarkable proportion of anticancer agents shows that it is important to continue to investigate natural products for their anticancer properties and to fuel anticancer drug development programs.

A group of natural products that have been shown to have anticancer properties as well as other interesting biological applications are the β-lactone containing group of natural products. Tetrahydroxylipstatin (THL, Orlistat) is a derivative of the natural product lipstatin, and is a β-lactone containing compound that is approved by the FDA as a pancreatic lipase inhibitor used to treat obesity[128,129]. Other members of this class include belactosin A, salinosporamide A and omuralide, which are all reported as potent inhibitors of the 20S proteasome (Figure 5.1)[130,131,132,133]. The ubiquitin-proteasome pathway plays a major role in eukaryotic cellular protein degradation[134]. This system adjusts the level of proteins involved in regulating cellular processes such as signal transduction, immune responses and cell cycle progression[135,136,137]. Inhibitors of the proteasome have gained attention for their ability to block cell cycle progression and cause apoptosis, which makes them useful antiproliferative agents[138]. Bortezomib and

carfilzomib are proteasome inhibitors that have been approved by the FDA for treatment of multiple myeloma, and several other proteasome inhibitors are currently in clinical trials[139,140].

In order to identify novel natural products with potential applications in human health, a library of bacteria were screened using the metabolomics approach outlined in Chapter 2. In this chapter, the isolation, strucrure elucidation and biological activity of two novel β-lactone containing natural products, cystargolides A and B is described.

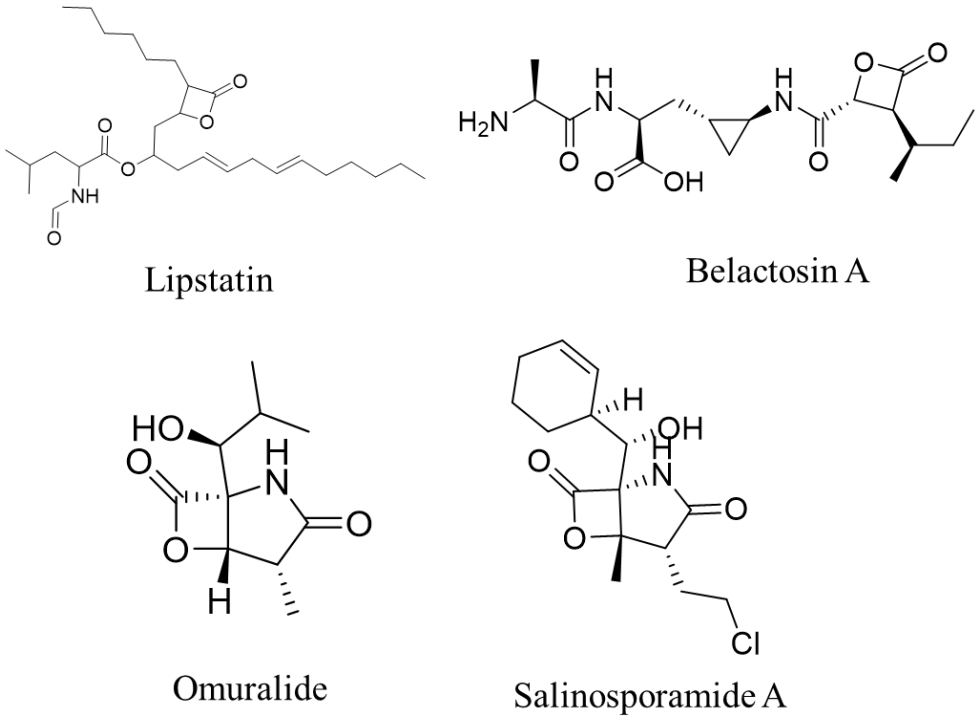


Figure 5.1 Molecular structures of β-lactone containing natural products.

5.2 MATERIALS AND METHODS

5.2.1. Large scale fermentation and extraction of *Kitasatospora cystarginea* and purification of cystargolides A and B

Kitasatospora cystarginea was fermented on a larger scale as described in Chapter 4.2.1. The 5 L fermentation was extracted as previously described with ethyl acetate and then partitioned between 80 % aqueous acetonitrile and hexanes. The acetonitrile phase was evaporated and analyzed for the presence of cystargolide A and B by LC-MS using the "chemical screening" method described previously (Chapter 2).

The acetonitrile fraction was separated by reverse phase flash chromatography using CombiFlash Rf (Teledyne ISCO) with a 15.5 g C_{18} column (High Performance GOLD, RediSep Rf), and eluted with H_2O (solvent A1) and MeOH (solvent B1) using a linear gradient from 50 to 100 % solvent B1 over 15 min followed by 100 % solvent B1 for an additional 5 minutes. A flow rate of 30 mL/min was used and eluent was detected by UV (214 nm). A mixture of cystargolide A and B eluted from 6.0-8.5 minutes.

Cystargolides were purified using a Thermo Surveyor HPLC system coupled with an evaporative light-scattering detector Sedex 55 with a Luna 110 Å phenyl hexyl column (5 µm, 250 x 10.00 mm, Phenomenex). A flow rate of 3 mL/min, using H_2O/0.1 % formic acid (solvent A2) and MeOH/0.1 % formic acid (solvent B2) was used to elute cystargolides A and B. The fraction containing cystargolides was separated using a linear gradient increasing from 50 % solvent B2 to 80 % solvent B2 over 17 minutes, followed by a ramp to 100 % solvent B2 over 2 minutes and then an isocratic wash with 100 % solvent B2 for an additional 12 minutes. Cystargolides A and B eluted at 12.1 and 10.2 minutes respectively, which were detected by ELSD and UV (220 and 254 nm). To

obtain enough material for full biological testing seven additional large scale fermentations and purifications were done using the same methods.

5.2.2 Structure elucidation of cystargolides A and B

Cystargolide A: Pale yellow power; $[\alpha]^{25}_{D}$ -18˚ (c=0.29, MeOH); IR ν_{max} 3294, 2962, 1835, 1719, 1646, 1534, 1465, 1201, 1025 cm^{-1}; ^{1}H and ^{13}C NMR see Table 5.1; (+) HRESIMS *m/z* 371.2177 $[M+H]^{+}$ (calcd for $C_{18}H_{31}N_2O_6$, 371.21766).

Cystargolide B: Pale yellow powder; $[\alpha]^{25}_{D}$ -28˚ (c=0.12, MeOH); IR ν_{max} 3299, 2964, 1838, 1723, 1648, 1529, 1467, 1157, 1024 cm^{-1}; ^{1}H and ^{13}C NMR see Table 5.2; (+) HRESIMS *m/z* 357.2021 $[M+H]^{+}$, (calcd for $C_{17}H_{29}N_2O_6$, 357.20201).

Optical rotations were measured on a Rudolph Autopol III polarimeter using a 50 mm microcell (1 mL). Infrared spectra were recorded using attenuated total reflectance, with samples deposited as a thin film on a Bruker Alpha FT-IR spectrometer. NMR spectra were obtained on a 600 MHz Bruker Avance III spectrometer equipped with a 1.7 mm inverse probe. Chemical shifts (δ) are reported in ppm and were referenced to the DMSO-d_6 residual peaks at δ_H 2.50 ppm and δ_C 39.51 ppm and coupling constants (*J*) are reported in Hz with the abbreviations (s) singlet, (d) doublet, (t) triplet, (q) quartet, (m) multiplet. LC-HRMS data were recorded used a Thermo Q Exactive fitted with ESI source, PDA, and LT-ELSD Sedex 80 detectors. High-resolution mass spectra were measured on a Thermo LTQ Orbitrap Velos mass spectrometer. Tandem MS analysis was conducted in positive mode by direct infusion of cystargolide A and B at a rate of 2 µL/min using an ESI source and collision-induced dissociation (CID) energy of 35eV.

5.2.3 Stereochemical analysis by Marfey's Method

Amino acid configurations were determined by Marfey's analysis of hydrolyzed cystargolides A and B[100]. To separate microconical vials, 20 µL of cystargolide A and B (10 mg/mL in MeOH) were added and then solvent was evaporated. HCl (250 µL of 6M) was added to each vial along with a stir bar and heated to 70-80 ºC for 75 minutes. Once the reaction mixtures had cooled, 1 mL of 1N $NaHCO_3$ followed by 20 µL of 1-fluoro-2,4-dinitrophenyl-L-alanine (L-FDAA) (10 mg/mL in acetone) was added to each reaction vial. The reactions were heated at 30-40 ºC for 1 hour before quenching with 100 µL of 6M HCl. The reaction mixture was reduced in volume under air, then diluted to 1 mL with 50:50 MeOH/H_2O for LC-MS analysis. 10 µL of derivatized amino acids were analyzed by LC-HRMS with a Hypersil Gold 100 Å column (Thermo, 1.9 µm C_{18} 50 x 2.1 mm). Compounds were eluted using a linear gradient from 95 % diH_2O/0.1 % formic acid (solvent A3) and 5 % acetonitrile/0.1 % formic acid (solvent B3) to 40 % solvent B3 over 55 minutes followed by a ramp to 100 % solvent B3 over 2 minutes then a hold for 3 minutes. A flow rate of 400 µL/min was used. Eluent was detected by ESI-MS monitoring *m*/*z* 120-800 in positive mode and UV (200-600 nm). Retention times were compared to that of authentic derivatized standards (Sigma Aldrich) to determine the amino acid configurations.

5.2.4 20S proteasome inhibition assay

Cystargolide A and B were evaluated for proteasome inhibition using purified human erythrocyte 20S proteasome (Enzo Life Sciences: BML-PW8720-0020). A solution of 20S proteasome was diluted to a final concentration of 3 µg/mL in assay buffer (50 mM Tris/HCl, pH 7.5, 25 mM KCl, 10 mM NaCl, 1mM $MgCl_2$, and 0.03%

SDS), and incubated with varying concentrations of inhibitors at 30°C for 10 minutes. To determine the chymotrypsin-like activity, the reaction was initiated by the addition of the fluorogenic substrate Suc-LLVY-AMC (Enzo Life Sciences, BML-9802-9090) at a final concentration of 75 μM. The rate of cleavage of the substrate was determined by measuring the fluorescence using a Spectra Max M2 (Molecular Devices) plate reader with an excitation wavelength of 360 nm and emission of 460 nm. The fluorescence was recorded every 15 seconds for 30 minutes, and the linear regression between 15 and 30 minutes were used to calculate the rate of substrate cleavage (AFU/s). Negative control wells were included that contained no inhibitor to show the maximum substrate cleavage rate, epoxomicin (0.5 μM) as a positive control and no enzyme (Blank) to show the minimum response. The concentration required to reduce the enzyme response by 50 %, (IC_{50}) were calculated by Prism 6.0 (GraphPad Software) using a nonlinear regression dose-response, variable slope model based on triplicate measurements ± standard deviation.

5.2.5 Human breast cancer and human fibroblast cell line cytotoxicity assays

Cystagolides A and B were evaluated for cytotoxicity against human breast adenocarcinoma cells (ATCC HTB-26) and human foreskin BJ fibroblast cells (ATCC CRL-2522). Breast adenocarcinoma cells were grown in Dulbecco's Modified Eagle's Medium/Nutrient Mixture F-12 Ham containing fetal bovine serum (10 %), penicillin (100 μU/mL) and streptomycin (0.1 mg/mL) with 5 % CO_2 at 37ºC. Human foreskin BJ fibroblast cells were cultured in Eagle's minimal essential medium containing fetal bovine serum (10 %), penicillin (100 μU/mL) and streptomycin (0.1 mg/mL) with 5 % CO_2 at 37ºC. Both cells lines were grown to 80 % confluency and diluted with their

respective growth medium without antibiotics. To a microwell plate, 90 µL of BJ fibroblast cells were added at a concentration of 1 x 10^4 cells/well and 90 µL of HTB-26 at a concentration of 5 x 10^3 cells/well and incubated with 5 % CO_2 at 37ºC for 24 hours. After 24 hours cells were treated with 10 µL of cystargolide A or B at final concentrations ranging from 1µg/mL to 128 µg/mL in sterile DMSO. Blank wells containing only growth medium and DMSO were used as positive controls, untreated wells containing growth medium, DMSO and cells were used as negative controls, and wells treated with either doxorubicin (HTB-26) or zinc pyrithione (fibroblast BJ) were used as positive treated controls. All samples were repeated in triplicate. Microwell plates were incubated with 5 % CO_2 at 37ºC for 72 hours (HTB-26) or 24 hours (fibroblast BJ). After incubation, alamarBlue® was added at a final concentration of 10 % v/v. Fluorescence was measured at 560/12 excitation and 590 nm emission at time zero, and four hours after addition of alamarBlue® using a Thermo Scientific Varioskan Flash plate reader. The change in fluorescence was used to calculate the percent viability of each treatment compared to negative control wells. The GI_{50} values, the concentration required to reduce the cell growth by 50 %, were calculated by Prism 6.0 (GraphPad Software) using a nonlinear regression dose-response, variable slope model based on triplicate measurements ± standard deviation.

5.2.6 Developmental Therapies Program Single Dose Cytotoxicity

Cystargolide B (10 mg) was sent to the National Cancer Institute (Bethesda, MD) for cytotoxicity evaluation against a panel of 60 cell lines at a single concentration of 10 µM. The assays were performed as described by the Developmental Therapies Program, Drug Discovery and Development Services[141]. GI_{50} values in the presence of cystargolide

B were calculated based on the reduction in cell viability compared to untreated negative controls.

5.3 RESULTS AND DISCUSSION

5.3.1 Large scale fermentation, extraction and isolation of cystargolides A and B

Using the metabolomics based chemical screening method described in Chapter 2, *K, cystarginea* was identified as the producer of unique metabolites based on the presence of two metabolites defined by the following two features: 3.0 min, *m*/*z* 371.2177 $[M+H]^+$) and 2.8 min, *m*/*z* 357.2021 $[M+H]^+$. Database searches (AntiBase 2012)[142] for the exact masses of these protonated adducts returned no matches, which led to a further chemical investigation of the *K. cystarginea* metabolome. A larger scale fermentation was undertaken, culture broths were extracted with ethyl acetate, and the resulting organic extract was fractionated by C_{18} liquid chromatography and further purified using RP-HPLC with a phenyl hexyl stationary phase to yield 2.5 mg and 3.1 mg of cystargolides A and B, respectively.

5.3.2 Structure elucidation and absolute configuration of cystargolides A and B

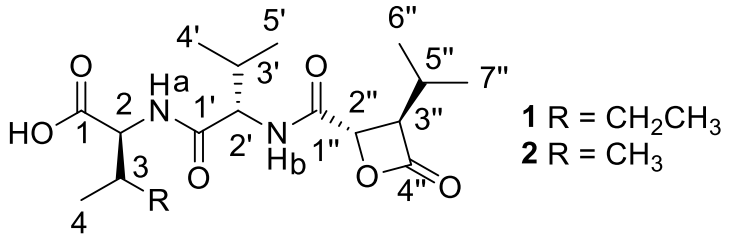
Cystargolide A was isolated as a pale yellow powder, and HRESIMS supported a molecular formula of $C_{18}H_{30}N_2O_6$ (*m*/*z* 371.2177 $[M+H]^+$, Δ=0.1 ppm), indicating five degrees of unsaturation. The NMR data (Table 5.1) revealed the presence of four carbonyls accounting for four of the five degrees of unsaturation in addition to two amide proton signals, which suggested that it was a peptide. Two amino acids were recognized as isoleucine and valine by interpretation of the ^1H-^1H COSY correlations identifying the two peptidic spin systems NH_a (δ_H 8.12) to H_3-6 (δ_H 0.82) and NH_b (δ_H 8.53) to H_3-5' (δ_H 0.81).

In a similar manner, COSY correlations assigned the third spin system H-2'' (δ_H 5.01) to H_3-7'' (δ_H 0.97). Key HMBC correlations between H-2'' / C-1'' (δ_C 167.3), C-4'' (δ_C 169.8); H-3'' (δ_H 3.50)/ C-1'', C-4'' and H-5'' (δ_H 2.11) / C-4'' unambiguously located the two carbonyl functional groups in position C-1'' and C-4'', respectively. The de-shielded resonances at δ_H 5.01 (H-2'') and δ_C 69.9 (C-2'') suggested that C-2'' was attached to an oxygen atom. The connectivity between the isoleucine, valine and the third spin system was determined based on the key HMBC correlations H-2 (δ_H 4.16), NH_a / C-1 (δ_C 172.8); H-2, NH_a, H-2' (δ_H 4.37) / C-1' (δ_C 170.5); H-2', NHb, H-2'', H-3'' / C-1''; and NH_b / C-2'' (δ_C 70.0). NOESY data supported the amino acid sequence assignment and showed strong correlations between NH_a (δ_H 8.12) / H-2' (δ_H 4.37); NH_b (δ_H 8.53) / H-2'' (δ_H 5.01), which was also supported by tandem mass spectrometry and the specific fragmentation pattern.

The fifth degree of unsaturation was attributed to cyclization. Two options were therefore available: a macrolactone cyclization between C-1 and the C-2'' hydroxy, or a β-lactone between C-4'' and the C-2'' hydroxy. The presence of IR absorption at 1835 cm^{-1} was characteristic of a β-lactone carbonyl stretching band and therefore strongly supported cyclization between C-4'' and the hydroxy at C-2''[143]. The lack of HMBC correlation between H-2'' and C-1 also supported a β-lactone cyclization. The presence of the β-lactone was further confirmed by the instability of cystargolide A in mild acidic conditions resulting in polymerization through condensation between two or three monomers[144]. To determine the relative configuration at the C-2'' and C-3'' positions on the β-lactone moiety, the $^3J_{HH}$ was used. A small $^3J_{H2''-H3''}$ of 3.8 Hz revealed that H-2'' and H-3'' have a *trans* configuration, thus the relative configuration of the ring was

determined to be (*2R*, 3S**)[145,146]. To the best of our knowledge, cystargolides are the first examples of β-lactone containing metabolites with a 3-isopropyl-4-oxo oxetate-2-carboxyl moiety. The molecular structure of cystargolides A and B is shown in Figure 5.2 as well as key NMR correlations used for structure elucidation. Chromatograms, MS, IR and NMR spectra for cystargolides A and B are shown in Appendix, Figures 8.31-8.52.

A



B

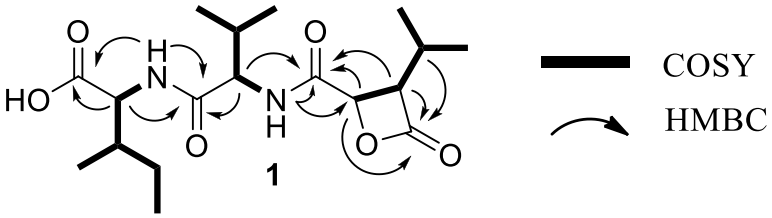


Figure 5.2. **A.** Molecular structure of cystargolide A (**1**) and B (**2**). **B.** Key NMR correlations used to determine the structure of cystargolide A.

Table 5.1. ^{1}H (600 MHz) and ^{13}C (150 MHz) NMR data for cystargolide A recorded in DMSO-d_6.

Position	δ_C, type	δ_H (J in Hz)	^{1}H-^{1}H COSY	HMBC (H→C)	NOESY
1	172.8, C				
2	56.4, CH	4.16, dd (7.1, 7.1)	3, NHa	1, 3, 4, 5, 1’	3, 4, NHa
3	36.0, CH	1.78, m	2, 4, 5	1, 2, 4, 5, 6	2
4	15.3,CH_3	0.86, m	3	3, 5	
5	24.5,CH_2	1.43, m	3, 6	2, 3, 4, 6	
		1.19, m	3, 6	2, 3, 4, 6	
6	10.9, CH_3	0.82, m	5	3, 5	
NHa		8.12, d (8.0)	2	1, 2, 3, 1’	2, 5, 2’, 3’
1’	170.5,C				
2’	57.2, CH	4.37, dd (8.0, 8.0)	3’, NHb	1’, 3’, 4’, 5’, 1’’	3’, NHa, NHb, 4’
3’	30.7, CH	2.01, m	2’, 4’, 5’	1’, 2’, 4’	2’
4’	18.8, CH_3	0.87, m	3’	2’, 3’, 5’	
5’	17.4, CH_3	0.81, m	3’	2’, 3’, 4’	
NHb		8.53, d (9.0)	2’	2’, 3’, 1’’, 2’’	2’, 3’, 2’’
1’’	167.3, C				
2’’	70.0, CH	5.01, d (4.0)	3’’	1’’, 3’’, 4’’, 5’’	NHb, 3’’, 5’’, 6’’
3’’	62.6, CH	3.50, dd (4.1,8.1)	2’’, 5’’	1’’, 2’’, 4’’, 5’’, 6’’, 7’’	5’, 2’’, 5’’, 7’’
4’’	169.8, C				
5’’	26.4, CH	2.11, m	3’’, 6’’, 7’’	2’’, 3’’, 4’’, 6’’, 7’’	2’’, 3’’
6’’	19.0, CH_3	1.00, d (6.7)	5’’	3’’, 5’’, 7’’	
7’’	19.2, CH_3	0.97, d (6.7)	5’’	3’’, 5’’, 6’’	

Table 5.2. ^{1}H (600 MHz) and ^{13}C (150 MHz) NMR data for cystargolide B in DMSO-d_6.

Position	δC, type	δH (J, Hz)	^{1}H-^{1}H COSY	HMBC (H→C)	NOESY
1	172.5, C				
2	57.2, CH	4.12 dd (6.6, 6.6)	3,NHa	1,3,4,1’	4
3	29.4, CH	2.06 m	2,4,5	1,2,5	
4	17.7, CH_3	0.89 m		5	
5	18.8, CH_3	0.88 m		2,3	
NHa		8.11 d (8.1)	2	1,3,1’,2’	5,2’,3’
1’	170.5, C				
2’	57.2, CH	4.38 t (7.7)	3’,NHb	5,3’,5’,1’’	3’,NHa,NHb,4’
3’	30.7, CH	2.01m	2’,4’,5’	5,2’,4’	
4’	18.9, CH_3	0.86 m	3’	2’,3’,5’	
5’	17.6, CH_3	0.82 d (6.7)	3’	2’,3’,4’	
NHb		8.53 d (9.0)	2’	2’,3’,1’’,2’’	2’,3’,2’’
1’’	167.0, C				
2’’	69.9, CH	5.01 d (3.8)	3’’	1’’,3’’,4’’, 5’’	3’’,5’’,NHb,6’’
3’’	62.5, CH	3.51 dd (3.9, 8.1)	2’’,5’’	1’’,2’’,4’’,5’’,6’’,7’’	2’’,5’’,6’’
4’’	169.7, C				
5’’	26.4, CH	2.12 m	3’’,6’’,7’’	2’’,3’’,4’’,6’’,7’’	3’’
6’’	19.2, CH_3	1.00 d (6.7)	5’’	3’’,4’’,5’’	
7’’	19.3, CH_3	0.97d (6.5)	5’’	3’’,4’’,5’’	

The absolute configurations of the amino acids in cystargolide A were determined using Marfey's analysis[100]. The peptides were hydrolyzed in HCl, derivatized using L-FDAA and compared to authentic derivatized amino acid standards using LC-HRMS (Appendix, Figures 8.48, 8.49). Retention times of derivatized standards were as follows: L-Val: 28.70 min, D-Val: 35.03 min, L-Ile: 34.02 min, D-Ile: 40.44 min, L-*allo*-Ile: 34.27 min and D-*allo*-Ile: 40.52 min. The derivatized hydrolysate of cystargolide A resulted in peaks that eluted at 28.8 and 34.0 minutes, which corresponded to L-Val and L-Ile respectively. Mosher's method was attempted to determine the absolute configuration of the β-lactone moiety after hydrolysis of cystargolides with 1N NaOH. Unfortunately, neither R nor S Mosher's derivatives were detected by LC-HRMS analysis which is most likely due to the instability of the β-lactone moiety leading to poly-condensation of cystargolide A and B.

Cystargolide B was isolated as a pale yellow powder, and HR-ESIMS supported a molecular formula of $C_{17}H_{28}N_2O_6$ (*m/z* 357.2021 $[M+H]^+$, Δ=0.2 ppm), which suggested the absence of one methylene group compared to cystargolide A. The NMR spectra of cystargolide B closely resembled those of cystargolide A while HSQC correlations revealed the absence of two signals, δ_H 1.43 (H_2-5a) and δ_H 1.19 (H_2-5b), which were in agreement with valine replacing isoleucine (Table 5.2). The interpretation of COSY, HMBC, and tandem mass spectrometry data further confirmed these observations and indicated the replacement of isoleucine with a second valine residue. Marfey's analysis indicated that both valine residues possessed an L-configuration (Appendix, Figures 8.47, 8.50) and the $^3J_{HH}$ coupling constants similarly revealed a *trans* substituted β-lactone.

Since the publication of the cystargolide A and B structures, another research group was able to complete a total synthesis, and determine the absolute configuration of the β-lactone substituents[147]. They were able to synthesize both diasteriomers of cystargolide A and B, and compare ^{1}H and ^{13}C NMR spectra to the natural products. Based on the proton shifts of the lactone substituents and the methyl carbon signals the absolute stereochemistry at the 2'' and 3'' positions were determined to be (*2R, 3S).*

5.3.3 Human 20S proteasome inhibition assay

The presence of a β-lactone ring in the structures of cystargolide A and B suggested that these natural products may be proteasome inhibitors, and so they were evaluated for their ability to inhibit human 20S proteasome in an enzyme assay. Varying concentrations of cystargolide A and B were incubated with purified human 20S proteasome. To determine the chymotrypsin-like activity, the fluorogenic substrate Suc-LLVY-AMC was added and fluorescence was recorded to measure the rate of enzyme activity. A dose-response curve was used to calculate IC_{50} values (Figure 5.3), which were determined to be 0.36 μM ± 0.017 and 0.93 μM ± 0.032 for A and B respectively.

Cystargolide A was observed to be a more potent inhibitor than B suggesting that the additional methylene results in a small change to the bulkiness of the peptide side chain that increases the interaction with the 20S proteasome. The ability of cystargolide A and B to inhibit chymotrypsin-like activity of the proteasome is not surprising given their similarity to the natural products belactosins A and C. The β-lactone containing belactosins A and C, inhibit chymotrypsin-like activity of rabbit 20S proteasome both with an IC_{50} value of 0.21 μM[131]. This is similar to the value observed for cystargolide A, and slightly more potent than B. Several medicinal chemistry and structure activity

relationship studies of the belactosins have resulted in synthetic analogues and hybrids with increased potency and cell permeability leading to improved compounds for the development of antiproliferative agents[148,149,150]. The clinically used agents carfilzomib and bortezomib inhibit the proteasome *in vitro* with an IC_{50} value of 5 nM and a K_i value of 0.6 nM respectively[139,140]. The great potency of these compounds lead to their development for clinical use, which suggests that the cystargolides may benefit from further structure optimization to increase their potency. The cystargolides represent excellent leads for continued development as inhibitors of the ubiquitin-proteasome pathway and cell proliferation.

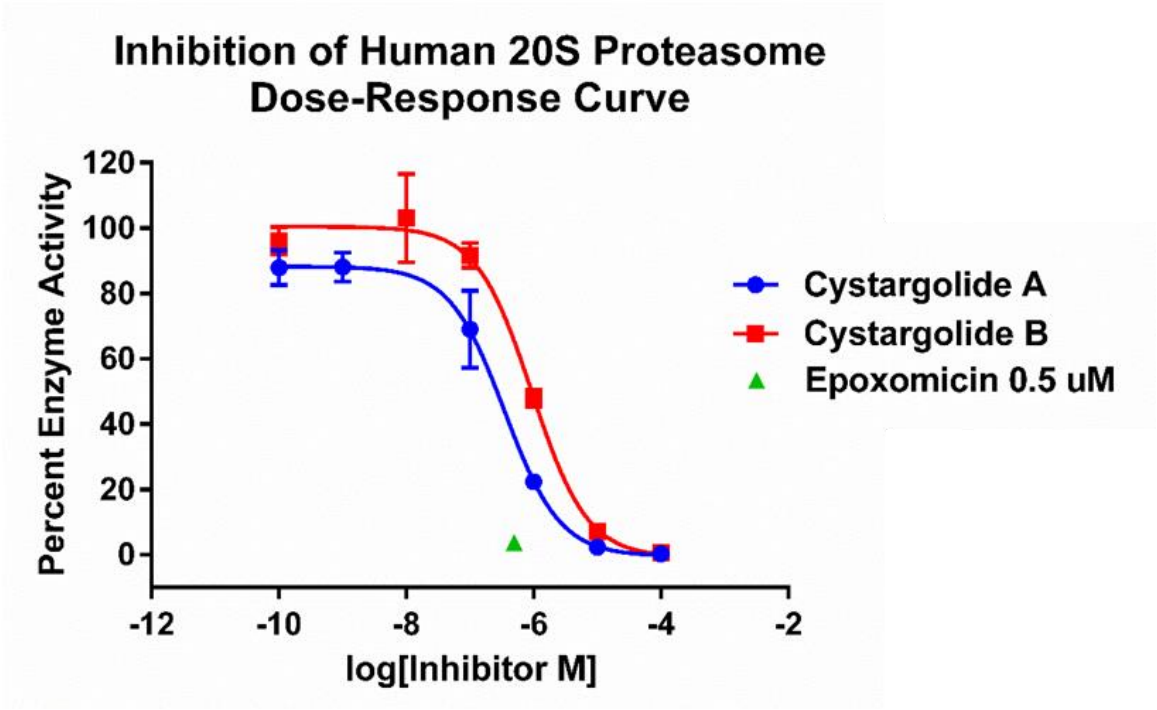


Figure 5.3. Dose-response curve showing the effects on the rate of Suc-LLVY-AMC cleavage by the human 20S proteasome (AFU/s Ex 360 nm, Em 460 nm) in the presence of increasing concentrations of cystargolides A and B and epoxomicin (0.5 µM) as a positive control.

5.3.4 Antiproliferating activity of cystargolides A and B

Since cystargolides A and B displayed potent inhibition of the validated molecular target for anticancer agents, the human 20S proteasome, they were evaluated for cytotoxicity in cell based assays. Cytotoxicity was determined against a cancer cell line, human breast adenocarcinoma cells and a healthy cell line, human foreskin fibroblast BJ cells (Figure 5.4). After incubating cells with varying concentrations of cystargolides A or B, the difference in fluorescence compared to untreated wells was calculated and used to determine percent cell viability. The results were plotted as a dose response curve, and GI_{50} values were determined to be greater than 250 μM. The high concentrations of cystargolides A and B required to reduce the growth of breast adenocarcinoma cells show that the potency observed in the *in vitro* enzyme based assay does not translate into potent *in vivo* cell based activity.

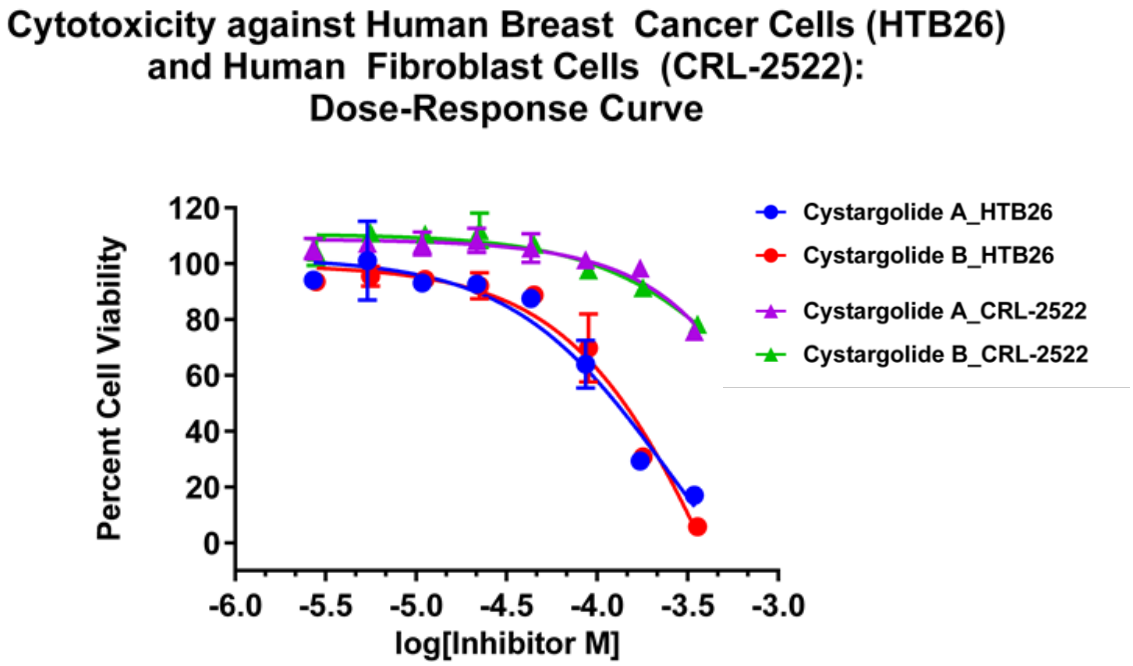


Figure 5.4. The percent cell viability of a human breast cancer cell line (HTB26) and a human fibroblast cell line (CRL-2522) in the presence of increasing concentrations of cystargolide A and B.

Cystargolide B was sent to the National Cancer Institute for evaluation in the Developmental Therapies Program 60-cell-one dose screen. Growth inhibition of each of the 60 cell lines in the presence of cystargolide B were measured and calculated as percent growth inhibition compared to untreated controls. The calculated percent growth inhibition against each cell line is shown in Figure 5.5. The mean growth percent observed at a single dose (10^{-5} M) of cystargolide B was 97.03 % with a range of 45.69. The cell lines that were the most susceptible to cystargolide B were colon cancer HT29, non-small cell lung cancer NCI-H522 and breast cancer T-47D, which had percent growths of 68.12 %, 70.38 %, and 74.12 % respectively. Significant growth inhibition at the single dose was not observed so cystargolide B was not selected as an anticancer candidate to be further evaluated in the NCI 5-dose screen.

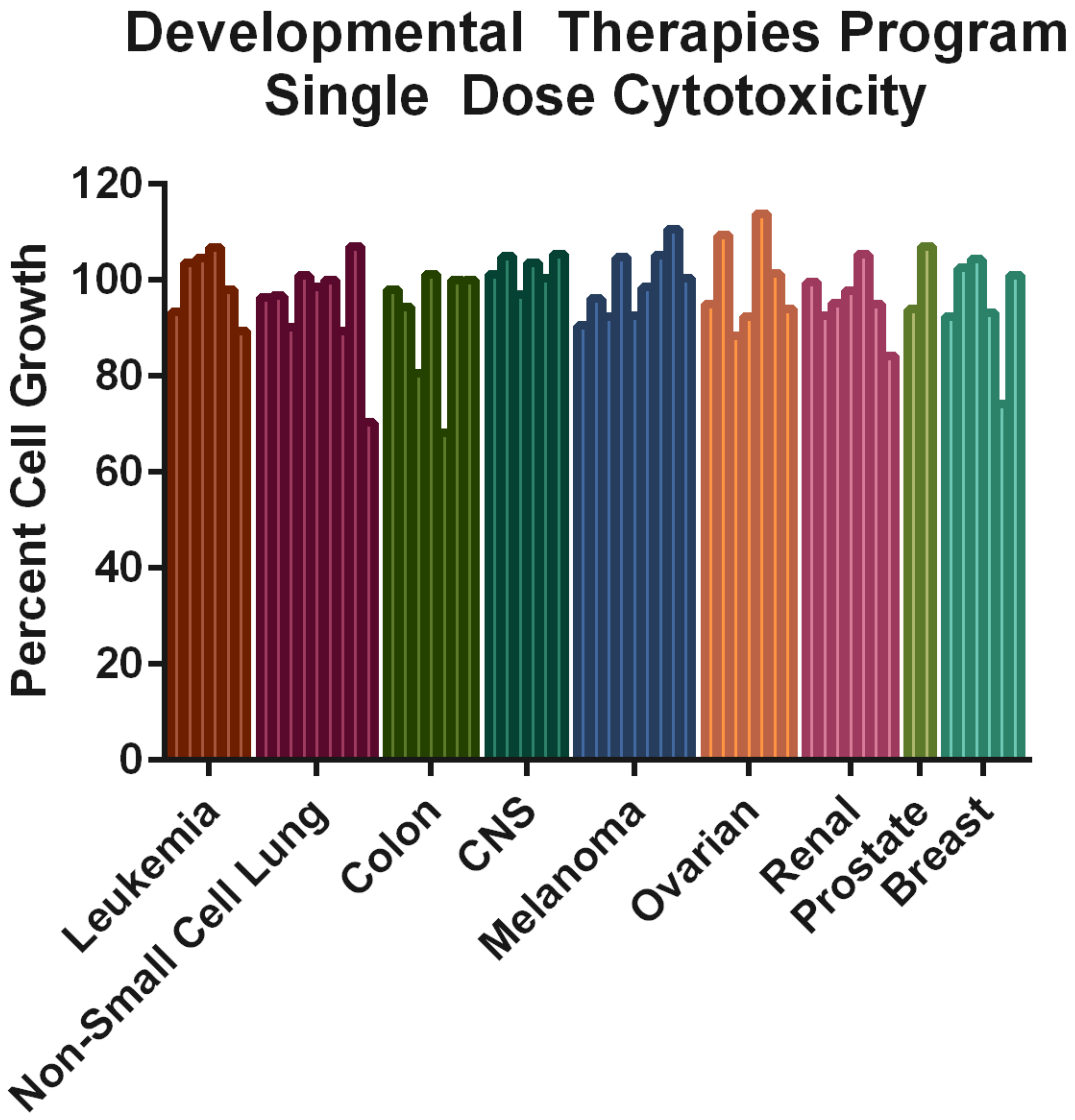


Figure 5.5. Antiproliferating activity of cystargolide B against the NCI-60 human tumor cell lines.

A significant decrease in potency was observed for *in vivo* cell based assays compared to that of an *in vitro* enzyme based assay. This decrease in potency could be caused by several reasons when a more complex whole cell system is used. Cystargolide B contains a relatively unstable β-lactone moiety that is required for activity, which could get hydrolyzed and therefore inactivated once it enters the cell. Cystargolide B may also be less potent because of off target interactions with other enzymes present in the cell. The lack of *in vivo* activity could also be a result of the inability to cross the cell membrane and enter the cell. Lack of cell permeability by cystargolides A and B would prevent the inhibitors from interacting with the proteasome and decrease their potency. Comparatively, the percent growth observed for breast adenocarcinoma cells HTB-26 is much lower than fibroblast BJ cells in the presence of high concentrations of cystargolides A and B. This shows that cystargolides A and B are selectively more toxic to cancer cells compared to normal cells and therefore have the potential for a large therapeutic window.

5.4 CONCLUSION

Two novel peptide natural products were isolated from the culture broth of the soil actinomycete *K. cystarginea.* These unique natural products contain a short, linear peptide chain coupled to an isopropyl and keto-substituted β-lactone group. To the best of our knowledge, these novel metabolites are the only natural products that have been described that contain this substitution pattern. This shows that cystargolide A and B were a unique discovery and contributes to the chemical space and biosynthetic capabilities of microorganisms.

Cystargolides A and B displayed potent inhibition of the human 20S proteasome, which is a validated molecular target for anticancer therapy. The potency of cystargolide A and B decreased when evaluated in a cell based assay, however they remained non-toxic to normal human fibroblast cells even at very high concentrations. The potentially large therapeutic window and potent *in vitro* enzyme inhibition of cystargolide A and B make them an attractive candidate for improvement using a medicinal chemistry approach. Generating a library of analogs focusing on greater stability of the β-lactone group, increased cell permeability and more specific proteasome interaction could yield a synthetic analog with an increase in *in vivo* activity and result in an excellent candidate for use as a clinical antiproliferating agent for the treatment of cancer.

The research group led by Dr. Tello-Aburto completed the total synthesis of cystargolides A and B. They were able to determine that the absolute stereochemistry was *2R, 3S* for both cystargolide A and B[147]. The completion of a total synthesis of cystargolides A and B also confirms the structure of these natural products.

CHAPTER 6: IDENTIFICATION OF THE BIOSYNTHETIC GENE CLUSTERS RESPONSIBLE FOR PRODUCING CYSTARGAMIDE AND CYSTARGOLIDES A AND B IN THE GENOME SEQUENCE OF *KITASATOSPORA CYSTARGINEA*

6.1 INTRODUCTION

Advances in molecular biology and sequencing techniques over the last two decades have allowed for an eruption of information about the genes and enzymes that are responsible for the biosynthesis of natural products. One class of natural products that has garnered a great deal of interest regarding their biosynthesis is nonribosomal peptides (NRPs). The biosynthesis of NRPs involves condensation of amino acids that is catalyzed by large multimodular enzymes, nonribosomal peptide synthetases (NRPSs)[151]. In addition to the 20 proteinogenic amino acid building blocks, NRPs commonly contain nonproteinogenic amino acids and can be linked to sugars, fatty acids, or polyketide groups[152]. Incorporation of these unusual amino acids and subsequent modifications results in a large amount of structural variety within this diverse group of peptide natural products.

NRPSs are large enzymes that catalyze the formation of peptide chains for NRPs[153]. In the canonical organization of these enzymes, they have a modular structure where each module is responsible for the addition of one amino acid in the growing peptide chain (Figure 6.1). The order of these modules is the same as the order in which each amino acid is incorporated and thus, determines the sequence of amino acids in the final structure. The genes that encode these modules are usually organized in the same order as the modules in the resulting enzyme. This is referred to as the colinearity rule. Each module contains three essential domains to catalyze the addition of a single amino acid, the adenylation domain (A), thiolation domain (T) and condensation domain (C). The A domain catalyzes the first step, and is responsible for activating the incoming amino acid by coupling of an adenylate group[154]. The A domain is highly specific, and

controls the identity of each monomer that gets incorporated. Each A domain contains a "specificity-conferring code" that consists of 10 amino acids that are involved in substrate binding[154]. The sequence of each code is specific for the amino acid that each A domain activates, and thus can be used to predict the identity of the amino acid incorporated by each A domain. This is very useful for predicting the amino acid sequence in the final natural product structure based on the gene sequence of adenylation domains. The thiolation domain (T) also referred to as peptidyl carrier protein (PCP) is responsible for tethering the activated amino acid to the NRPS enzyme[155]. The T domain contains a phosphopantethein arm that attaches to the activated amino acid via a thioester bond. The flexibility of the phosphopantethein arm allows for the tethered amino acid to extend and interact with adjacent domains that are involved in chain elongation and peptide modification. Lastly, the condensation domain is responsible for catalyzing the condensation reaction between the free amine of the incoming amino acid, and the growing peptide chain[151].

Other domains may be present within modules that are responsible for alterations to the peptide[156]. Epimerization (E) domains catalyze the epimerization of incoming amino acids from L to D, methyltransferase (MT) domains often catalyze the N-methylation of amino acids, and cyclization (Cy) domains catalyze nucleophilic attack of thiol or hydroxyl side chains of cysteine, threonine or serine on the adjacent amide carbonyl, followed by dehydration to give five membered thiazoline or oxazoline rings. The thioesterase domain (TE) is usually the final domain at the terminal end of the NRPS, which catalyzed the release of the natural product from the enzyme by cleavage of the thioester bond either through hydration or macrocyclization[156].

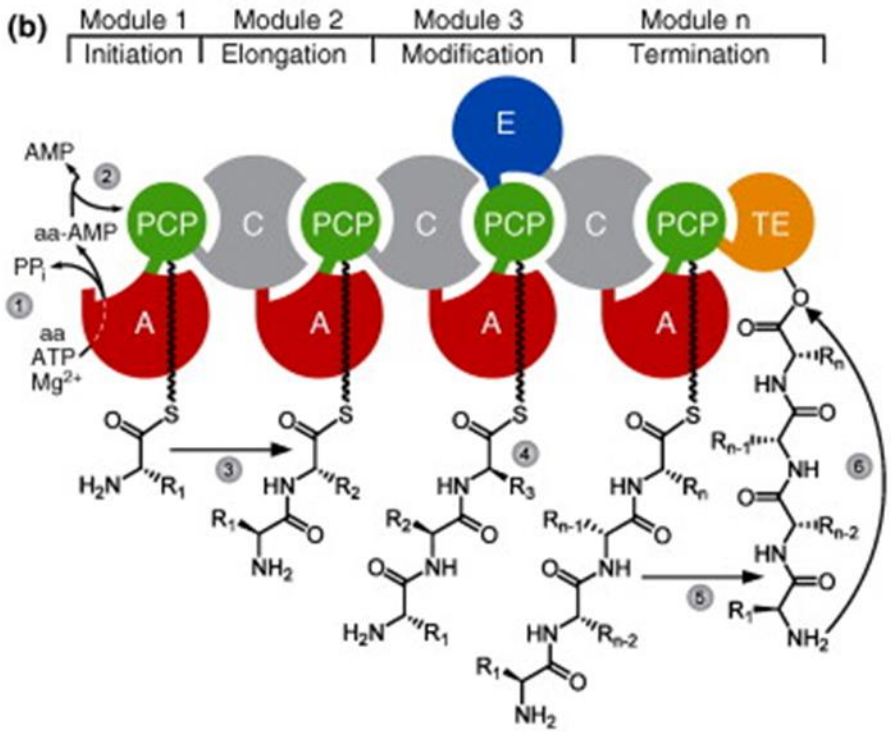


Figure 6.1. General mechanism for biosynthesis of nonribosomal peptides by Nonribosomal Peptide Synthetase enzymes (M. Strieker *et. al.)*[152]

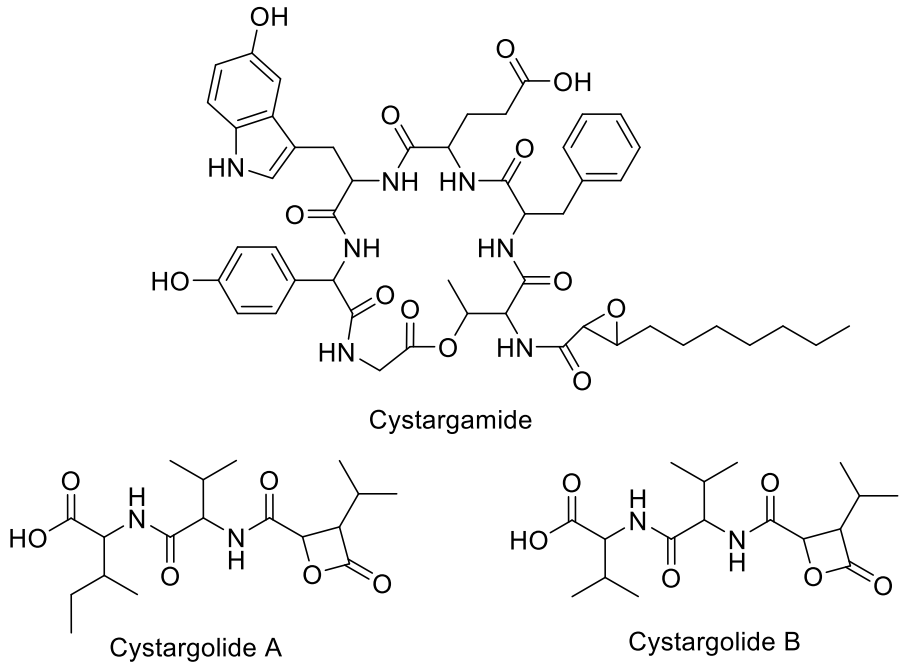


Figure 6.2. Molecular structures of cystargamide, cystargolide a and cystarolige B

In addition to modifications, NRPs can be hybridized with other types of natural products like polyketides (PKs), fatty acids (lipopeptides) and sugars (glycopeptides). The genes responsible for coupling these groups to the peptide chain are often located in close proximity to the NRPS genes, or may be 'borrowed' from primary metabolism and encoded elsewhere on the genome sequence[153].

As described in preceding chapters, the peptide natural products cystargamide and cystargolides A and B were isolated from the fermentation broth of *Kitasatospora cystarginea* (Figure 6.2). The biosynthetic route for both of these types of natural products was predicted to involve NRPS enzymes due to the peptide chains in both structures. Cystargamide is a lipopeptide that contains non-proteinogenic amino acids and is coupled to a fatty acid group, with a similar structure to the calcium dependent antibiotic (CDA). It was hypothesized that the biosynthetic gene cluster responsible for producing cystargamide would have similar features to the gene cluster that produces CDA, and thus contain genes to encode NRPSs. The cystargolides contain a short linear peptide chain that is coupled to a very unusual β-lactone containing subunit. It was unclear what the origin of the β-lactone containing subunit is, but we hypothesized that the biosynthetic gene cluster would contain NRPS domains to synthesize the dipeptide chains, and link them to the acyl substituted β-lactone. To determine the gene clusters involved in biosynthesis of these novel natural products, the whole genome sequence of *K. cystarginea* was sequenced, and evaluated for the presence of putative biosynthetic gene clusters. Herein we describe the genome sequence, annotation, and identification of the gene clusters responsible for production of cystargamide and cystargolides A and B.

6.2 MATERIALS AND METHODS

6.2.1 Genome sequencing, assembly and annotation

K. cystarginea NRRL B-16505 was obtained from the Agriculture Research Services Culture Collection under the accession number NRRL B-16505. To isolate genomic DNA (gDNA), 10 mL of ISP2[49] broth was inoculated with 10 µL of a *K. cystarginea* spore stock (10^9 cfu/mL) and fermented at 30 ºC for 4 days. High molecular weight genomic DNA was isolated using the Qiagen genomic tip 100/G DNA isolation kit and assessed for appropriate size and purity using agarose gel electrophoresis. Long insert DNA library preparation and Pacific Biosciences RS II sequencing was performed by McGill University and the Genome Quebec Innovation Centre. Four single molecule real-time (SMRT) cells were used. Genome assembly was performed using the hierarchical genome assembly process 2 by McGill University and the Genome Quebec Innovation Centre[157].

Coding sequences and gene function were predicted using Rapid Annotation using Subsystems Technology 2.0 (RAST)[158-160]. AntiSMASH 3.0.5 was used to annotate the secondary metabolite biosynthetic gene clusters present in the genome sequence of *K. cystarginea*[161].

6.2.2 Construction of a cystargolide gene disruption mutant

The CRISPR/cas9 genome editing plasmid, pCRISP0myces-2 was used to create gene disruption mutants[162]. The pCRISPomyces-2 plasmid was obtained from Addgene (Addgene plasmid # 61737), where it was deposited by Huimin Zhao. Genomic DNA from *K. cystarginea* was isolated using a phenol/chloroform/isoamyl alcohol extraction method[163] for use as template DNA for PCR reactions. PCR primers and DNA

oligonucleotides used in this study are listed in Table 6.1. The cystargolide (CyO) spacer insert was prepared by annealing 5 µM of the oligonucleotides, CyOG2PS_F and CyOG2PS_R in HEPES buffer (30 mM, pH 7.8), to a total volume of 100 µL. The oligonucleotide mixture was heated to 95 ºC for 5 min, then cooled to 4 ºC at a rate of 0.1 ºC/sec. The annealed spacer was cloned into pCRISPomyces-2 using the Golden Gate Assembly method[164]. pCRISPomyces-2 (100 ng) was combined with 0.3 µL of annealed insert, 2 µL of ligase buffer (10X), 1 µL of T4 ligase (NEB), 1 µL of BbsI (NEB), and sterile H_2O to a final volume of 20 µL. A thermocycler (Eppendorf Mastercycler) was used the facilitate assembly with the following conditions: Step 1: 37 ºC for 10 minutes and 16 ºC for 10 minutes (repeated for 10 cycles), step 2: 50 ºC for 5 min, step 3: 65 ºC for 20 minutes, and step 4: 4 ºC for storage. The resulting ligation reaction was used to transform chemically competent *E. coli* NEB 5α cells, and transformants were selected on lysogeny broth (LB) agar plates (10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl, 20 g/L agar) containing apramycin (Apr) (50 µg/mL). Isopropyl β-D-1-thiogalactopyranoside (IPTG) (100 µM) and 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) (20 µg/mL) were also included in the medium to facilitate identification of recombinant colonies via blue white screening. Recombinant colonies were cultured in LB broth containg Apr (50 µg/mL) at 37 ºC for 24h, and plasmid DNA was extracted using a plasmid isolation kit (Qiagen) to give pCRISPCyOPS.

Two homologous arms (0.85 kb) flanking the cystargolide protospacer region were amplified by PCR using the primer pairs CyOG2_FL / CyOG2_RL and CyOG2_FR / CyOG2_RR with the conditions listed in Table 6.2. Resulting amplicons were gel-purified. A three way Gibson assembly was used to clone both the left and right

homologous template arms into pCRISPCyOPS. XbaI (NEB) and Shrimp Alkaline Phosphatase (rSAP) (NEB) were used to digest and dephosphorylate pCRISPCyOPS prior to assembly. To assemble the homologous template arms into pCRISPCyOPS, 23 ng of each homologous template arm was mixed with 100 ng of linear pCRISPCyOPS in 2X Gibson Assembly Master Mix (NEB) with a final reaction volume of 20 µL. Reactions were incubated at 37 ºC for 60 min followed by 65 ºC for 10 min. The ligation reaction mixture was used to transform chemically competent *E. coli* NEB 5α cells. Transformants were selected on LB agar containing Apr (50 µg/mL). A total of 48 colonies were patched onto Apr containing LB agar plates, and after 24 hours of growth, were screened for the homologous template insert using colony PCR with DMSO extracted template DNA and CyOG2_FL and CyOG2_RR PCR primers. PCR conditions used are listed in Table 6.2. Transformants yielding a 1.7 kb amplicon were inoculated into 10 mL of LB broth containing Apr (50 µg/mL) and cultured at 37 ºC for 24 hours with agitation at 200 rpm. pCRISPCyOPSHT plasmid DNA was extracted and correct assembly was confirmed using sequencing performed by Eurofins MWG Operon (Huntsville AL, USA) using the primers BbsI_Fwd, BbsI_Rev, XbaI_Fwd, XbaI_Rev, CyOG2_1.2_SF and CyOG2_1.2_SR.

Table 6.1. Oligonucleotides used for PCR, assembly of pCRISPCyOPSHT and sequencing.
Homologous regions are shown in capital letters, non-homologous regions in lower case letters and sticky ends are shown as underlined text.

Primer Name	Primer Sequence (5'→ 3')
CyOG2_PS_F	ACGCACAGCGGTACGACGTCGGGA
CyOG2_PS_R	AAACTCCCGACGTCGTACCGCTGT
CyOG2_FL	gccgggcgttttttatctaGACCGCGCGGCGTGGACC
CyOG2_RL	ccgcagggTCTTCGCCTACCCGCGCGG
CyOG2_FR	ggcgaagaCCCTGCGGGTCGAGCCGG
CyOG2_RR	tacggttcctggcctctagaCGGGAGCACCGGCGCGGT
BbsI_Fwd	TGTGAATGGCCTGTTCGG
BbsI_Rev	CGAGCGTTCTGAACAAATCC
XbaI_Fwd	CAAGTTAAAATAAGGCTAGTCC
XbaI_Rev	GTCCTGTCGGGTTTCGC
CyOG2_1.2_SF	GGGAACGCGATCCGGTGC
CyOG2_1.2_SR	GGTACCGGTACTGCTGTCG
CyOG2_0.2_SF	GCAGGGCGGTGAAGGTGC
CyOG2_0.2_SR	ACGCTGCTCCCGGTGACC
16S_27F	AGAGTTTGATCMTGGCTCAG
16S_1525R	AAGGAGGTGWTCCARCC
16S_936R	GGGGTTATGCCTGAGCAGTTTG

Table 6.2. PCR conditions showing primer pairs, temperatures used for each step, number of cycles and polymerase enzyme that was used.
Each step was 30 seconds in duration, and each reaction had an initial 2 min denaturation step and a final elongation step for 4 min.

Forward Primer (5 µM)	Reverse Primer (5 µM)	Denaturation Temp (°C)	Annealing Temp (°C)	Elongation Temp (°C)	# Cycles	Enzyme
CyOG2_FL	CyOG2_RL	95	66	72	30	Econotaq® Plus Green (2X)
CyOG2_FR	CyOG2_RR	95	66	72	30	Econotaq® Plus Green (2X)
CyOG1_FL	CyOG2_RR	95	66	72	30	Econotaq® Plus Green (2X)
CyOG2_1.2_SF	CyOG2_1.2_SR	95	62	72	30	Econotaq® Plus Green (2X)
CyOG2_0.2_SF	CyOG2_0.2_SR	95	67	72	30	Econotaq® Plus Green (2X)
16S_27F	16S_1525R	95	54	72	30	Econotaq® Plus Green (2X)

pCRISPCyOPSHT was used to transform chemically competent *E. coli* ET12567::pUZ8002 and transformants were selected on LB agar plates containing Apr (50 µg/mL), kanamycin (Kan) (50 µg/mL) and chloramphenicol (Cml) (25 µg/mL). pCRISPCyOPSHT was introduced into *K. cystarginea* via conjugal DNA transfer. *E. coli* ET12567::pUZ8002::pCRISPCyOPSHT was inoculated in reduced salt LB (1 g/L of NaCl) containing Apr, Cml and Kan, and incubated at 37 ºC to an OD_{600} of 0.4-0.6, then rinsed twice with fresh reduced salt LB, and resuspended in 10 % of the culture volume with 2XYT (16 g/L of Bacto Tryptone, 10 g/L of yeast extract, 5 g/L of NaCl, pH 7.0). An aliquot of 50 µL of *K. cystarginea* spores (10^9 cfu/mL) were rinsed with fresh 2XYT and mixed with *E. coli* ET12567::pUZ8002::pCRISPCyOPSHT. Conjugation mixtures were plated onto MS agar (20 g/L mannitol, 20 g/L Baker's soya flour, 20 g/L agar, tap water) supplemented with 10 mM $MgCl_2$ and incubated at 30 ºC or 37 ºC. After 18 hours, plates were overlaid with 1 mL of sterile H_2O containing naladixic acid (Nal) (25 µg/mL) and Apr (50 µg/mL). Plates were incubated for an additional 7 days at 30 ºC. Exconjugants were patched to ISP2 agar plates.

6.2.3 Analysis of wild type and mutant strains

Exconjugants were screened for the production of cystargolides A and B by inoculating 10 mL of seed medium (10 g/L of glucose and 10 g/L of yeast extract) with a single colony from ISP2 agar plates. Seed cultures were fermented at 30 ºC and 200 rpm shaking for three days. 10 mL of lean production medium (1045) (Chapter 2), was inoculated with 500 µL of seed cultures and fermented for four days with the same growth conditions. Fermentation broths were extracted with 10 mL of ethyl acetate, and analyzed by LC-MS using the Kerr lab "chemical screening" method described

previously (Chapter 2). Spores of a positive mutant were preserved in 25 % glycerol and stored at -80 ºC. Fermentations were repeated for 12 replicates of the positive mutant by plating spores onto ISP2, and inoculating broth using a single colony. LC-MS analysis was also repeated using the same conditions.

PCR was used to amplify regions surrounding the predicted mutation. Template DNA was extracted using phenol/cholorform/isoamyl alcohol DNA extraction method, from 5 day cultures in ISP2 that were inoculated directly with mutant or wild type strains. PCR was performed using the CyOG2_1.2_SF / CyOG2_1.2_SR and CyOG2_0.2_SF / CyOG2_0.2_SR primer pairs, listed in Table 6.1 using the PCR conditions listed in Table 6.2. To confirm the taxonomic identity of the mutant and wild type strains, the 16S rRNA gene was amplified using PCR (27F and 1525R primers) and sequenced by Eurofins MWG Operon (Huntsville AL, USA) using the 936R primer[165].

6.3 RESULTS AND DISCUSSION

6.3.1 Genome sequence, assembly and annotation

In order to identify the biosynthetic gene clusters responsible for producing cystargolides A and B and cystargamide, the whole genome sequence of *K. cystarginea* was determined using Pacific Biosciences RS II sequencing. The sequencing efforts resulted in a total of 313 760 subreads that passed quality control with an average read length of 4 707 bp resulting in 164X coverage (assuming a 9 Mb genome size). The genome sequence was assembled and consisted of one major contig (8 967 537 bp) and 9 smaller contigs ranging in size from 286 393 to 813 bp. The total consensus sequence consisted of 9 351 686 bp and had a G+C content of 71.3 %.

Annotation using the RAST online interface[158-160] identified a total of 8 334 predicted coding sequences (CDSs) spanning 432 subsystems. The number of features for each subsystem is summarized in Table 6.3. The closest type strain neighbors of *K. cystarginea* based on the RAST genome annotation were *Streptomyces griseus* subsp. *griseus* NBRC 13350, *Streptomyces avermitillis* MA-4680, and *Streptomyces roseosporus* NRRL 11379.

AntiSMASH was used to predict the secondary metabolite biosynthetic gene clusters (BGCs) present in the genome sequence of *K. cystarginea.* AntiSMASH predicted a total of 36 BGCs including four NRPSs, four PKSs (type I and II), three terpenes, two bacteriocins, eight lantipeptides, one siderophore, nine hybrids, one melanin, one butyrolactone and three uncharacterized gene clusters, which are summarized in Table 6.4.

Table 6.3. Summary of subsystem features in the genome sequence of *Kitasatospora cystarginea*, predicted using RAST 2.0.

Subsystem Category	Subsystem Feature Counts
Cofactors, vitamins, prosthetic groups, pigments	317
Cell wall and capsule	203
Virulence, disease and defense	72
Potassium metabolism	27
Miscellaneous	41
Phages, prophages, transposable elements	4
Membrane transport	106
Iron acquisition and metabolism	57
RNA metabolism	127
Nucleosides and nucleotides	137
Protein metabolism	376
Cell division and cell cycle	37
Motility and chemotaxis	4
Regulation and cell signaling	55
Secondary metabolism	25
DNA metabolism	135
Fatty acids, lipids and isoprenoids	284
Nitrogen metabolism	26
Dormancy and sporulation	18
Respiration	151
Stress response	193
Metabolism of aromatic compounds	47
Amino acids and derivatives	639
Sulfur metabolism	43
Phosporus metabolism	37
Carbohydrates	507

Table 6.4. Summary of predicted biosynthetic gene clusters present in the genome sequence of *Kitasatospora cystarginea* using AntiSMASH 3.0.5.

Cluster	Type	Similar Known Gene Clusters (Percent of genes that show similarity)
1	Other	
2	T1 PKS	
3	TransAT PKS-NRPS	
4	Lantipeptide	
5	Bacteriocin	
6	NRPS	
7	TransAT PKS	Migrastain / dorrigocin (100 %)
8	T2 PKS-oliogiosaccharide-NRPS	
9	T2 PKS	Spore pigment (83 %)
10	NRPS	
11	NRPS	
12	Hybrid	Esmeraldin (52 %)
13	Lantipeptide	
14	NRPS-T1 PKS	
15	Lantipeptide	
16	Bacteriocin	
17	Siderophore	
18	T1 PKS	
19	Butyrolactone	
20	Melanin	
21	Lantipeptide	
22	Lantipeptide	
23	Terpene	
24	Lantipeptide	
25	Other	
26	Lantipeptide	
27	Terpene	Hopene (61 %)
28	Terpene	
29	Siderophore-T3 PKS	Desferrioxamine B (100 %)
30	Lantipeptide	
31	Other	
32	NRPS-Lantipeptide	
33	NRPS-T1PKS-Other	
34	NRPS	
35	Thiopeptide-Lantipeptide	
36	NRPS-T1 PKS	

Upon review of the predicted BGCs, the genes responsible for producing several known compounds were identified in the genome sequence of *K. cystarginea* (Figure 6.3). One of the minor metabolites that was purified from fermentation extracts of *K. cystarginea* was migrastatin (discussed in chapter 2). Not surprisingly, one of the BGCs that was predicted by AntiSMASH shared sequence homology with the gene cluster responsible for producing migrastatin, isomigrastatin and dorrigocin from *Streptomyces platensis* NRRL 18993[166]. All of the eleven genes present in cluster 7 showed similarity to genes responsible for producing migrastatin. The identification of this gene cluster within the genome of *K. cystarginea* strengthened the support that 16505_490 is migrastatin, and contributed to annotation of the genome sequence.

In addition to the migrastatin gene cluster, two other clusters shared homology to known BGCs. All of the six genes present in cluster 29 showed similarity to the genes responsible for producing the siderophore natural product desferrioxamine B in *Streptomyces coelicolor* M145[167]. Of the thirteen genes present in cluster 27, 61 % were homologous to genes responsible for the terpene natural product, hopene in *Streptomyces coelicolor* A3(2)[39,168]. These metabolites were not detected in any of the fermentation extracts of *K. cystarginea*, either because they were not produced in the culture conditions that were used, or they were not extracted and detected. The presence of these gene clusters show that this strain has the potential to biosynthesize these natural products, and they may be produced when *K. cystarginea* is fermented using different methods. There are also several BGCs that were predicted by AntiSMASH that show very little similarity to known gene clusters. This shows that *K. cystarginea* has the genetic potential to produce many unique natural products that have yet to be identified.

This is similar to what has been discovered by the genome sequencing efforts focused on the closely related genus, *Streptomyces.* A similar number of putative biosynthetic gene clusters are present in the genome sequences of many *Streptomyces* spp., which shows that *Kitasatospora* spp. have an equal genetic ability to produce natural products. The large number of unique BGCs in the *K. cystarginea* genome shows that members of this genus represent an excellent resource that have largely been underexplored for the identification of novel natural products.

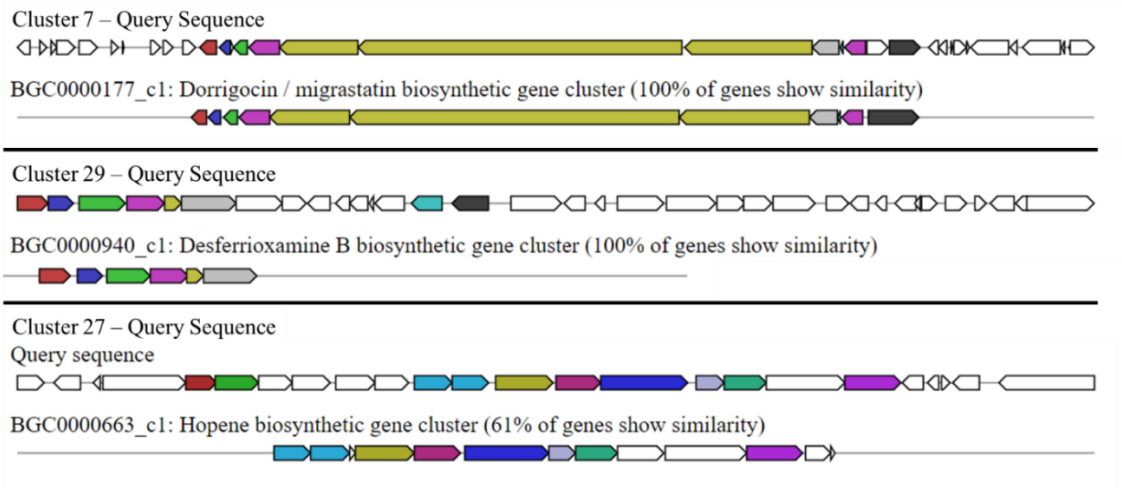


Figure 6.3. Alignment of biosynthetic gene clusters of known natural products to clusters present in the genome sequence of *Kitasatospora cystarginea* using AntiSMASH.

6.3.2 Bioinformatics analysis of putative cystargamide gene cluster

The peptidic structure of cystargamide suggested that it is biosynthesized using NRPSs, so the four NRPS and nine hybrid BGCs predicted by AntiSMASH were analyzed further for their potential role in the biosynthesis of this natural product. Cluster 11 was putatively identified as the gene cluster responsible for producing the lipopeptide cystargamide based on the number of NRPS modules and predicted substrates of the A domains (Figure 6.4). Details of each gene present in this gene cluster are summarized in Table 6.5. Cystargamide is a cyclic lipopeptide that contains 2,3-epoxy decanoic acid attached to the N-terminus of a peptide with the following sequence: Thr-Phe-Glu-Htrp-Hpg-Gly. Within cluster 11, ORF1989 encodes a NRPS that has five modules, each containing C, A and T domains. ORF1990 also encodes an NRPS that has a single module containing a C, A, T and TE domains. The amino acid sequences of the six adenylation domains in ORF1989 and ORF1990 were used to predict the amino acids that are incorporated at each elongation step using NRPSpredictor2[169]. Assuming the colinearity rule, the predicted amino acid sequence closely matched the structure of cystargamide. NRPSpredictor2 results are summarized in Table 6.6.

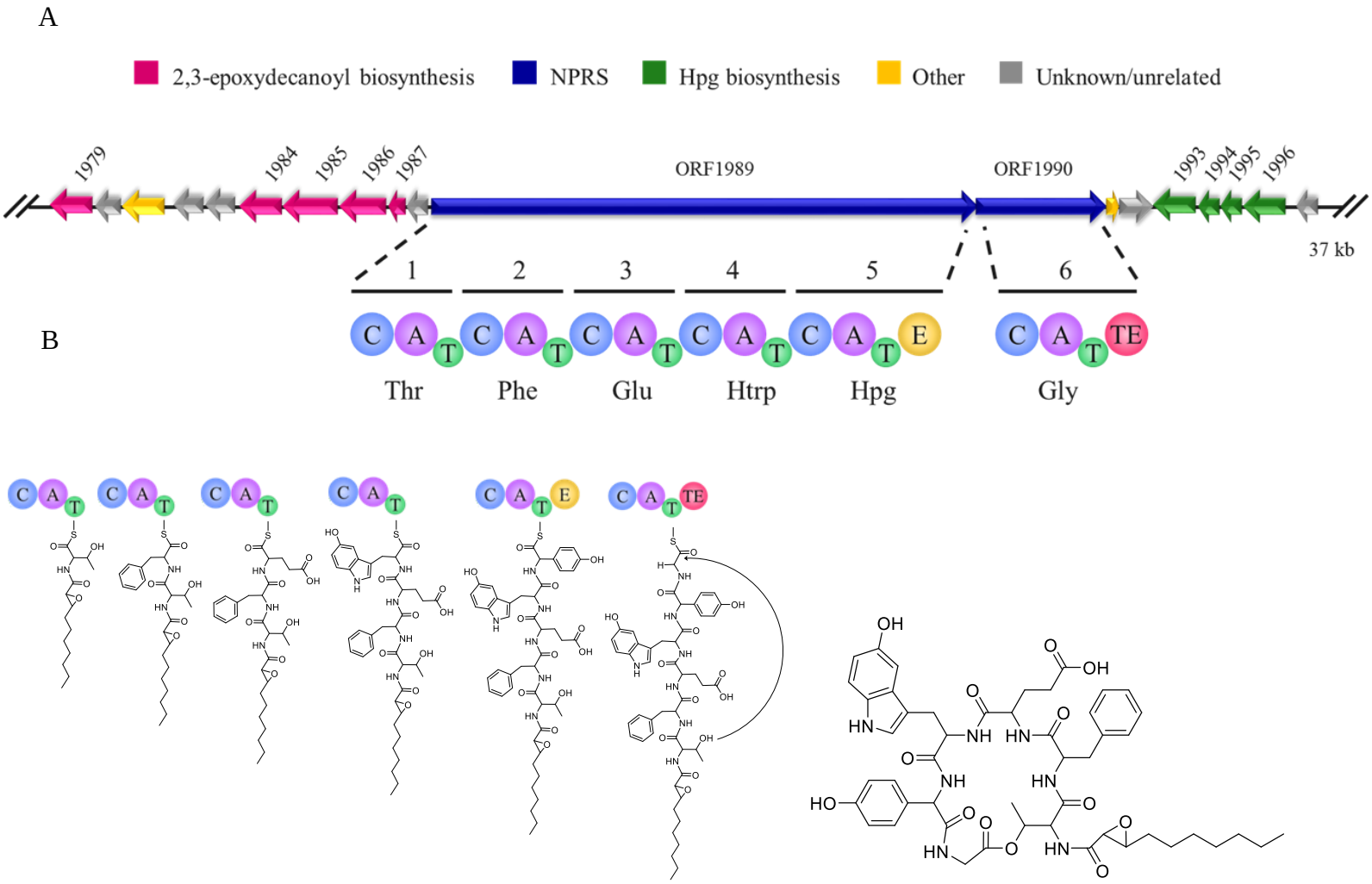


Figure 6.4. (A) Schematic representation of the 37 kb DNA region spanning cluster 11, the gene cluster putatively responsible for the biosynthesis of cystargamide.(B) Proposed mechanism of cystargamide biosynthesis.

Table 6.5. Summary of the bioinformatics analysis of the putative cystargamide gene cluster, Cluster 11 and their predictor role in biosynthesis.

ORF No.	Protein Length	Nearest Homologue: Description[a]	Nearest Homologue: % Identity, accession, origin[a]	Nearest Homologue: Accession[a]	Putative Function in cystargamide biosynthesis
1975	479	Multidrug transporter	81%, *Streptomyces durhamensis*	WP_051926938.1	Transport
1976	392	Hypothetical protein	80%, *Streptomyces durhamensis*	WP_031168828.1	
1977	333	Hypothetical protein	80%, *Streptomyces durhamensis*	WP_031168830.1	
1978	281	Alpha/beta hydrolase	37%, *Streptomyces afghaniensis*	WP_020275753.1	
1979	491	Probable long chain fatty acid CoA ligase PpuA	36%, *Pseudomonas putida*	AAM75414.1	Decanoyl-CoA ligase
1980	228	Type 12 methyltransferase	59%, *Streptomyces thermolilacinus*	WP_023588193.1	
1981	309	LysR family transcriptional regulator	51%, *Herbidospora daliensis*	WP_062435602.1	Transcriptional regulator
1982	217	ATP-dependent Clp protease proteolytic subunit	83%, *Streptomyces* sp. NRRL S-384	WP_030906597.1	
1983	206	ATP-dependent Clp protease proteolytic subunit	85%, *Streptomyces griseoplanus*	WP_055588379.1	
1984	342	3-oxoacyl-ACP synthase	66%, *Streptomyces* sp. TOR3209	WP_019331099.1	β-ketoacyl synthase III
1985	580	Acyl-CoA oxidase	66%, *Streptomyces* sp. NRRL S-455	WP_051704900.1	Decanoyl-CoA oxidase
1986	405	3-oxoacyl-ACP synthase	66%, *Streptomyces* sp. AcH 505	WP_041998922.1	β-ketoacyl synthase II
1987	81	Acyl carrier protein	48%, *Streptomyces coelicolor*	WP_011028857.1	Acyl carrier protein
1988	54	No significant similarity			

ORF No.	Protein Length	Nearest Homologue: Description[a]	Nearest Homologue: % Identity, accession, origin[a]	Nearest Homologue: Accession[a]	Putative Function in cystargamide biosynthesis
1989	5750	Non-ribosomal peptide synthetase	50%, *Actinoplanes friuliensis,*	WP_023362360.1	NRPS modules 1-5
1990	1304	Non-ribosomal peptide synthetase	56%, *Kibdelosporangium* sp. MJ126-NF4	CEL16414.1	NRPS module 6
1991	73	MbtH-Like Protein	79%, Multi *Streptomyces*	WP_015613088.1	Unknown-A domain activating enzyme[170]
1992	448	Hypothetical protein	49%, *Kibdelosporangium* sp. MJ126-NF4	WP_052478142.1	
1993	368	α-hydroxy acid oxidizing enzyme	78%, *Streptomyces canus*	WP_020123018.1	4-hydroxymandelate oxidase
1994	207	4-hydroxyphenylpyruvate dioxygenase	80%, *Streptomyces canus*	WP_020123017.1	4-hydroxyphenylpyruvate dioxygenase
1995	164	4-hydroxyphenylpyruvate dioxygenase	65%, *Streptomyces canus*	WP_020123017.1	4-hydroxyphenylpyruvate dioxygenase
1996	498	GntR family transcriptional regulator	69%, *Streptomyces lydicus*	WP_033269185.1	4-hydroxyphenylglycine aminotransferase
1997	109	No significant similarity			
1998	399	4-hydroxybenzoate-3-monooxygenase	83%, *Saccharothrix* sp. ST-888	WP_045302347.1	
1999	293	Carbon-nitrogen hydrolase	72%, *Streptomyces achromogenes*	WP_052423740.1	
2000	159	Hypothetical protein	64%, *Streptomyces xanthophaeus*	WP_031143167.1	

[a]Results generated by BLASTp analysis of non-redundant protein sequences (nr)

Table 6.6. The predicted amino acid that is activated by each adenylation domain using NRPSpredictor2.

ORF	Adenylation Domain Sequence	NRPSpredictor 2 Nearest Neighbor	Score (Percent)
1989	1	Threonine	90
1989	2	Phenylalanine	90
1989	3	Asparagine	50
1989	4	Phenylalanine	60
1989	5	Hydroxyphenylglycine	80
1990	6	Glycine	90

Four of the six predicted amino acids have high scores of 80 or 90 %. The identity of these amino acids that are predicted with the highest confidence match the location and identity in the observed structure of cystargamide. The two amino acids with lower confidence levels of 50 and 60 % do not match the structure of cystargamide. However, both predicted amino acids are similar to the amino acids present in the structure of cystargamide at those locations. NRPSpredictor 2 predicted a second phenylalanine at position four, which as a hydrophobic aromatic amino acid, is in the same category as hydroxytryptophan. The rarity of hydroxytryptophan in the structure of natural products makes it difficult for bioinformatics programs to predict its identity. Therefore since an amino acid similar to Htrp was predicted, it can be suggested that this adenylation domain is responsible for the incorporation of Htrp at this position. At position 3, NRPSpredictor2 predicted the incorporation of asparagine with 50 % confidence. Since this predicted amino acid is in the same hydrophilic category as glutamic acid, and the confidence level is relatively low, it is reasonable to suggest that this A domain is responsible for incorporating glutamic acid instead of asparagine.

In the fifth module encoded by ORF1989, there is an E domain in addition to the C, A and T domains. This suggests incorporation of a D-amino acid at this position. This predicted stereochemistry contradicts the structure of cystargamide, which contains L-Hpg at the fifth position, as determined by Marfey's method (Chapter 4). This suggests that the E domain present in ORF1989 is inactive. E domains are approximately 450 amino acids in length and contain seven highly conserved core motifs (E1-E7) that are essential for catalytic activity[110]. The amino acid sequence of ORF1989 contains six of the seven conserved motifs (E2-E7) but lacks conserved motif E1 (PIQxWF). It can

therefore be deduced from the stereochemistry observed in the structure of cystargamide, and the lack of an E1 motif that the E-domain present in ORF1989 is inactive.

ORF1990 encodes a NRPS that contains C, A, T and TE domains. The thioesterase domain is responsible for termination and cleavage of the peptide chain from the NRPS. The cyclization of macrolactone peptides is often catalyzed by the TE domain, where cleavage and cyclization occur in a single step, using a threonine or serine hydroxyl group for intramolecular nucleophilic attack[171]. This is likely the mechanism for cystargamide biosynthesis, since the TE domain is present in the final module that incorporates glycine, which is cyclized with the threonine side chain in the final structure of cystargamide.

This gene cluster also possesses several genes that are likely involved in lipoinitiation during the biosynthesis of cystargamide. ORF1987 encodes a protein with amino acid sequence homology to an acyl carrier protein (ACP), ORF1986 has sequence homology to 3-oxoacyl-ACP synthase (or β-ketoacyl-ACP synthase II, KASII) and ORF1984 is homologous to 3-oxoacy-ACP synthase (or β-ketoacyl-ACP synthase III, KASIII). These three enzymes are involved in the biosynthesis of fatty acids and are likely involved in the biosynthesis of decanoyl-ACP. Other enzymes involved in the biosynthesis of decanoic acid may be borrowed from primary metabolism since they are not present in close proximity to the rest of the cystargamide gene cluster. Enzymes involved in biosynthesis of fatty acid portions of lipopeptides are often absent from BGCs and it is common for microbes to recruit enzymes from primary metabolism[124]. Calcium dependent antibiotic (CDA) contains a 2,3-epoxyhexanoic acid group, and contains the same gene organization for fatty acid biosynthesis compared to cluster 11.

The CDA gene cluster contains genes encoding ACP, KASII and KASIII enzymes, and is also missing the genes encoding acetyl and malonyl transacylases, β-ketoacyl reductase, β-hydroxylacyl dehydratase and enoyl reductase, which are respsonsible for reduction of β-carbonyls during fatty acid biosynthesis[124].

Also present in cluster 11 are genes that encode enzymes involved in β-hydroxylation of fatty acids. ORF1979 encodes an enzyme that is homologous to putative fatty acid CoA ligases. Fatty acid CoA ligases are responsible for catalyzing the addition of coenzyme-A to free fatty acids, which is the first step in β-hydroxylation of fatty acids in the degradation pathway[172]. ORF1985 is homologous to acyl-CoA oxidase and shares 65 % sequence identity to *hxcO* (SCO3247) present in the calcium dependent antibiotic gene cluster in the genome of *S. coelicolor* A3(2)[39]. Acyl-CoA oxidases are also involved in the second step of hydroxylation of fatty acids, and are responsible for catalyzing the unsaturation of acyl-CoAs to *trans*-enoyl-CoAs. The presence of these enzymes suggest a possible mechanism for 2,3-epoxydecanoic acid may involve CoA linked intermediates (Figure 6.5). This mechanism would involve biosynthesis of decanoyl-ACP, followed by cleavage from the ACP using a fatty acid thioesterase (TE) to give free decanoic acid. Decanoic acid can then be coupled to CoA, which is desaturated by acyl-CoA oxidase to produce *trans*-decenoyl-CoA. A flavin dependent monooxygenase could then be responsible for epoxidation of the *trans*-decenoyl-CoA resulting in 2,3-epoxydecanoyl-CoA, which would then get coupled to amino acids to produce cystargamide. Cluster 11 does not contain a fatty acid thioesterase or any monooxygenase enzymes that are directly upstream or downstream of the fatty acid biosynthesis genes. However adjacent to the 4-hydroxyphenylglycine cassette is ORF1998, which shares homology with a gene

encoding a monooxygenase and therefore this gene may be involved in epoxidation. The mechanism involved in the biosynthesis of the epoxy-fatty acid group in CDA differs from the typical mechanism for β-oxidation of fatty acids. The biosynthetic gene cluster of CDA contains an acyl-CoA oxidase (*hxcO)* and a flavin dependent monooxygenase (*hcmO)* involved in the biosynthesis of the epoxy-fatty acid group[124]. Despite the sequence similarity of HxcO and HcmO to enzymes that act on CoA linked intermediates, the enzymes preferencially act on ACP-linked intermediates[173]. It was also shown that HxcO possesses intrinsic enoyl-ACP epoxidase activity and is sufficient to produce the epoxide ring alone, and that HcmO may be involved in an alternate route. The similarity between the genes present in the CDA gene cluster and cluster 11 suggests that the 2,3-epoxydecanoyl group in cystargamide may be biosynthesized in a similar manner to the 2,3-epoxyhexanoyl group in CDA (Figure 6.5). However the cystargamide gene cluster contains a fatty acid CoA ligase, which is lacking in the CDA gene cluster. At this stage, either mechanism is a possibility for the biosynthesis of 2,3-epoxydecanoic acid and it is not clear whether the epoxide ring is formed from a CoA linked intermediate, like the typical mechanism of β-oxidation of fatty acids, or from an ACP linked intermediate, like CDA.

This gene cluster also contains genes responsible for biosynthesis of the non-proteinogenic amino acid 4-hydroxyphenylglycine. The biosynthesis of this amino acid involves a cassette that encodes three enzymes that are used to transform 4-hydroxyphenylpyruvate to 4-hydroxyphenylglycine (Figure 6.6)[174]. The first step involves conversion of 4-hydroyphenylpyruvate to L-4-hydroxyphenylmandelate by 4-hydroxyphenylpyruvate dioxygenase (HmaS). Cluster 11 contains two genes, ORF1995

and ORF1994 with amino acid sequence homology to HmaS. The amino acid sequence of ORF1994 shares 62 % identity with the verified HmaS from *Amycolatopsis orientalis* (O52791.1), while ORF1995 shares 43 % sequence identity. This shows that ORF1994 is likely the active enzyme responsible for the first step of 4-hydroxyphenylglycine biosynthesis in *K. cystarginea.* The next step in biosynthesis is oxidation of L-4-hydroxyphenylmandelate to give 4-hydroxybenzoylformate, which is catalyzed by hydroxymandelate oxidase (Hmo)[174]. Hmo is similar to FMN containing glycolate oxidase and mandelate dehydrogenase (MDH) and shares 50 and 52 % sequence identity respectively. ORF1993 has similar amino acid sequences to this family of oxidizing enzymes, with 78 % sequence identity to the α-hydroxyacid oxidizing enzyme from *Streptomyces canus*, 65 % sequence identity to hydroxymandelate oxidase from *Streptomyces* sp. Tue 6075 and 58 % sequence identity to the validated Hmo from *Amycolatopsis* orientalis (O52792.1). This shows that ORF1993 is likely responsible for the oxidation of 4-hydroxyphenylmandelate to 4-hydroxybenzylformate, the second step in 4-hydroxyphenylglycine biosynthesis. The last biosynthetic step involves the conversion of 4-hydroxybenzoylformate to 4-hydroxyphenylglycine using hydroxyphenylglycine transaminase (HpgT)[174]. ORF1996 displays amino acid sequence homology to the GntR family of transcription regulators, which are classified as pyridoxal phosphate (PLP)-dependent aspartate aminotransferases. HpgT is also a member of the PLP-dependent aminotransferases and uses L-Tyr as an amino donor. This step results in 4-hydroxyphenylglycine, and regeneration of 4-hydroxyphenylpyruvate from L-Tyr, which feeds into a new cycle of 4-hydroxyphenylglycine biosynthesis[174]. ORF1996 shares 56 % sequence identity with the amino acid sequence of HpgT from

Amycolatopsis orientalis (O52815.1). This reveals that ORF1996 is likely responsible for the final biosynthetic step to generate 4-hydroxyphenylglycine.

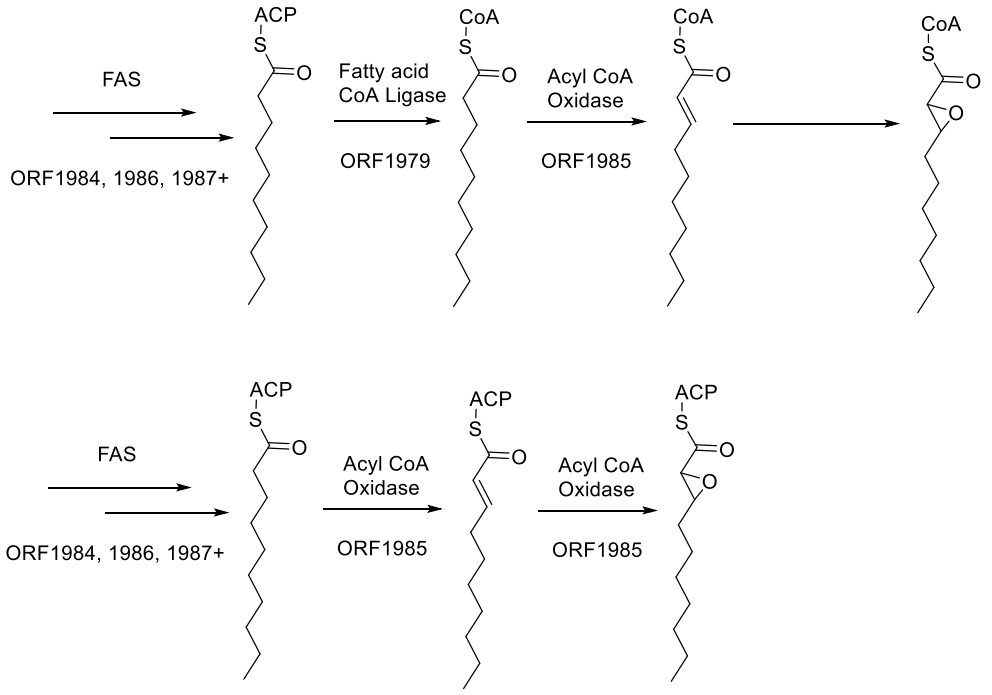


Figure 6.5. Possible mechanism for the biosynthesis of 2.3-epoxydecanoic acid, an intermediate in the biosynthesis of cystargamide.

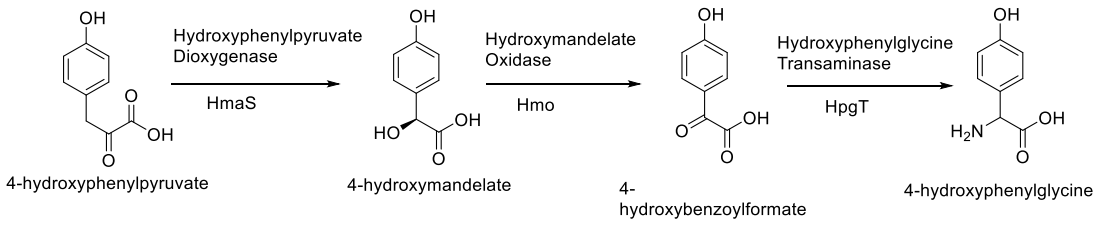


Figure 6.6. Biosynthesis of the non-proteinogenic amino acid 4-hydroxyphenylglycine (Adapted from Hubbard et. al.[174]).

The structure of cystargamide contains an additional non-proteinogenic amino acid, 5 hydroxytryptophan (Htrp). Incorporation of this amino acid into microbial natural products is much rarer than Hpg, and its biosynthetic origin is unclear[175]. In humans, 5-hydroxytryptophan is an intermediate in serotonin biosynthesis and is generated by hydroxylation of L-tryptophan using the enzyme tryptophan hydroxylase (EC 1.14.16.4)[176]. The method of biosynthesis of Htrp in bacteria is yet to be determined. There are no tryptophan hydroxylase homologs present within close proximity to the cystargamide cluster suggesting that the gene responsible for hydroxylation of tryptophan is located somewhere else in the genome sequence of *K. cystarginea.* There are several genes that are annotated as putative hydroxylases that are present in the genome sequence of *K. cystarginea* that may be responsible for the hydroxylation of tryptophan.

In summary, cluster 11 is a gene cluster that was predicted using AntiSMASH and is likely responsible for the production of cystargamide. This gene cluster contains two multidomain NPRSs that are predicted to produce a peptide with an amino acid sequence that is very similar to the sequence of cystargamide. Cluster 11 also contains the gene cassette required for biosynthesis of the non-proteinogenic amino acid, Hpg, which is present in the structure of cystargamide. In addition, cluster 11 also contains several genes involved in the biosynthesis and epoxidation of fatty acids. To experimentally confirm the role of cluster 11 in the biosynthesis of cystargamide, deletion of the entire gene cluster was attempted using CRISPR/cas9 mediated multiplex genome editing but these attempts were unsuccessful due to the difficulty of conjugal DNA transfer to *K. cystarginea.* However, the overwhelming evidence determined by bioinformatics analysis

supporting that cluster 11 is responsible for the production of cystargamide rendered additional mutation efforts unnecessary.

6.3.3 Bioinformatics analysis of putative cystargolide gene cluster

Given the unique structural characteristics of cystargolide A and B, identifying the putative gene cluster responsible for production was more challenging than cystargamide using bioinformatics predictions. Because of the linear peptide chains that are present in the structures of cystargolides A and B, it was hypothesized that NRPS enzymes may be involved in the biosynthesis. The NPRS and hybrid gene clusters that were predicted by AntiSMASH, were evaluated for their potential role, and one cluster was identified as the most likely candidate. Cluster 34 was selected because it contained the appropriate number of NRPS modules, three A, and T domains and four C domains.

Cluster 34 contains five ORFs that encode NRPS-like domains involved in peptide biosynthesis (Figure 6.7). All ORFs in cluster 34, and their predicted functions are summarized Table 6.7. ORF7573 encodes a stand-alone C domain, ORF7572 encodes a single module containing a C, A, and T domain, ORF7570 and ORF7569 each encode one A and one T domain, and ORF7568 encodes a C and T domain. NRPS enzymes are traditionally classified as large, multi-modular enzymes that contain several domains encoded by the same gene[153]. Each multi-modular NRPS is typically responsible for catalyzing several steps in the biosynthesis of peptide natural products. Cluster 34 diverges from this usual NRPS gene organization in that several genes encode mono- or di-domain enzymes. This noncanonical NRPS organization is also observed in the biosynthetic gene clusters of the siderophore natural products, enterobactin and vibriobactin[177]. H. Wang and co-workers conducted a comprehensive study to mine all of

the available complete genome sequences, spanning three domains of life, for the presence of NRPS, PKS and hybrid gene clusters[178]. They determined that out of 1 428 NPRS gene clusters identified from bacterial genome sequences, 18 % were organized in a non-modular arrangement. They proposed that this non-modular NRPS gene organization is more common than originally anticipated and thus, should be re-classified as type-II since it resembles the organization of type-II PKSs. The non-modular arrangement of the cystargolide biosynthetic gene cluster shows that it falls within this proposed group of type-II NRPSs.

NRPSpredictor2 and Minowa prediction were used to predict the amino acid selectivity of each A domain (Table 6.8). The A domain encoded by ORF7569 (A3) was predicted using both methods to incorporate Val with 70 % confidence score. No consensus could be reached for the remaining two A domains and predictions were made with low confidence scores. However, The A domain in ORF7570 (A2) was predicted to incorporate a hydrophobic amino acid by both prediction methods and thus may be responsible for incorporation of either Ile or Val for cystargolides A and B respectively. The low confidence scores using prediction methods is likely due to the promiscuity of this A domain, and shows that it may have the ability to activate multiple different amino acids. No consenses was reached for the prediction of the amino acid incorporated by the A domain in ORF7572 (A1). The commonality between the two prediction methods was that the unit is likely hydrophilic in nature. The low confidence score could be a result of incorporation of an unusual starter or extender unit[179]. This aligns with the observed structure of cystargolides A and B since they contain an unusual isopropyl and carboxy substituted β-lactone group.

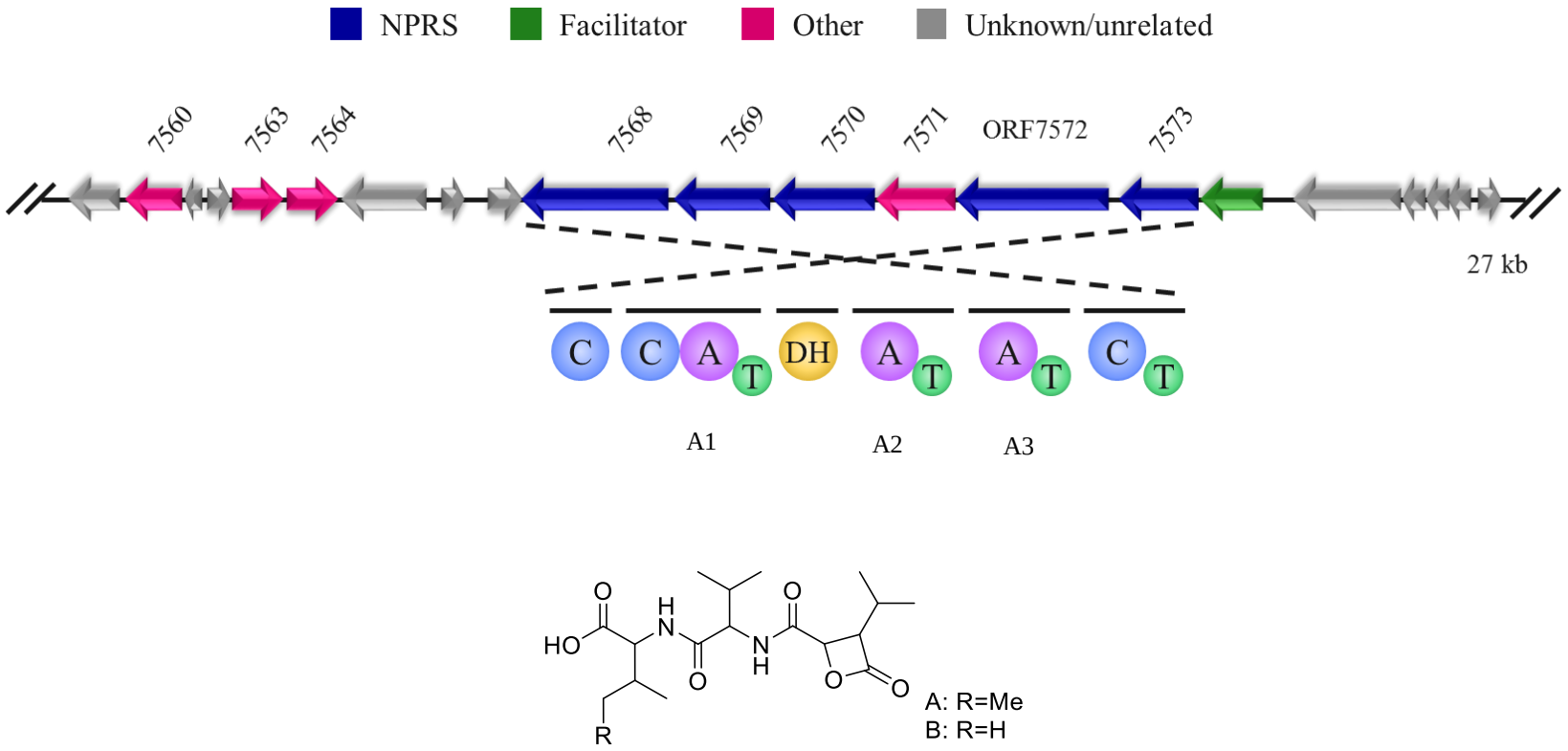


Figure 6.7. Schematic representation of the 27 kb DNA region spanning cluster 34, the gene cluster involved in the biosynthesis of cystargolide A and B.

Table 6.7. Summary of the bioinformatics analysis of cluster 34, the gene cluster involved in biosynthesis of cystargolides A and B.

ORF No.	Protein Length	Nearest Homologue: Description[a]	Nearest Homologue: Sequence identity, origin[a]	Nearest Homologue: Accession[a]
7557	121	Hypothetical protein	75%, *Streptomyces rubellomurinus*	WP_045699647.1
7558	510	NAD(P)-dependent oxidoreductase	84%, *Kitasatospora azatica*	WP_035840390.1
7559	321	Ribokinase	65%, *Streptacidiphilus albus*	WP_034086908.1
7560	263	Enoyl-CoA hydratase	76%, *Streptomyces* sp. WM6372	WP_053694337.1
7561	43	Hypothetical protein-YkuD superfamily	65%, *Streptomyces albus* subsp. Albus	KUJ67803.1
7562	70	Cupin	70%, *Streptomyces* sp. W007	WP_032791300.1
7563	270	Methyltransferase	72%, *Streptomyces hygroscopicus*	WP_060953975.1
7564	248	Short-chain dehydrogenase	84%, Multi *Kitasatospora*	WP_057235080.1
7565	516	Hypothetical protein-Peptidases_S53 superfamily	69%, Multi *Streptomyces*	WP_043389144.1
7566	149	Hypothetical protein AQJ27 49940	62%, *Streptomyces olivochromogenes*	KUN33597.1
7567	278	Hypothetical protein-aliginate lyase superfamily	74%, *Saccharothrix* sp. ST-888	WP_052681550.1
7568	946	NRPS	83%, *Saccharothrix* sp. ST-888	WP_045299923.1
7569	599	NRPS	80%, *Saccharothrix* sp. ST-888	WP_045299925.1
7570	612	NRPS/PKS hybrid	85%, *Saccharothrix* sp. ST-888	WP_045299927.1

ORF No.	Protein Length	Nearest Homologue: Description[a]	Nearest Homologue: Sequence identity, origin[a]	Nearest Homologue: Accession[a]
7571	390	Acyl-CoA dehydrogenase	96%, *Saccharothrix* sp. ST-888	WP_045299929.1
7572	1039	NRPS	85%, *Saccharothrix* sp. ST-888	WP_045299931.1
7573	423	Hypothetical protein-Condensation Supefamily	83%, *Saccharothrix* sp. ST-888	WP_045299987.1
7574	462	MFS transporter	77%, *Kitasatospora* sp. MY 5-36	WP_049659092.1
7575	707	Mannosyltransferase	51%, *Streptomyces* sp. 769	WP_052287171.1
7576	42	No significant similarity		
7577	119	Hypothetical protein	73%, *Streptomyces caatingaensis*	WP_049716363.1
7578	129	Hypothetical protein	91%, *Saccharothrix* sp. ST-888	WP_045304688.1

[a]Results generated by BLASTp analysis of non-redundant protein sequences (nr)

Table 6.8. The predicted amino acid that is activated by each adenylation domain using NRPSpredictor2 and Minowa prediction.

ORF	Adenylation Domain Sequence	NRPSpredictor 2: Nearest Neighbor	Score (Percent)	NRPSpredictor2: Three clusters	Minowa prediction
7572	1	Thr	50%	Hydrophilic	Lys
7570	2	Phe	50%	--	Leu
7569	3	Val	70%	Hydrophobic-aliphatic	Val

In addition to incorporation of the proteinogenic amino acids valine and isoleucine, cystargolides A and B also contain the very unusual isopropyl and carboxy-substituted β-lactone moiety. To the best of our knowledge, there are no other known natural products that contain a β-lactone with this substitution pattern. The most similar structures to cystargolide A and B are the belactosins, however their biosynthesis remains undetermined[130]. The biosynthesis of macrolactones in NRPs and PKs has largely been investigated and as a result is a well characterized step in natural product biosynthesis[153]. Macrocyclization is often the final step in biosynthesis and involves intramolecular nucleophilic attack of a hydroxyl on the enzyme bound carbonyl and is often catalyzed by a terminal thioesterase domain. This results in cleavage of the thioester for release from the enzyme and cyclization in a single step. However the formation of 4-membered lactone rings is less extensively studied and consequently the mechanisms and genes involved are not yet fully understood. The biosynthetic gene clusters involved in the formation of the β-lactone containing natural products, ebelactone[180], salinosporamide A[181], lipstatin[182] and oxazolomycin[183] have been investigated and resulted in the proposal of different mechanisms for β-lactone formation. The proposed mechanism of β-lactone biosynthesis in salinosporamide A does not involve the thioesterase that is present in the gene cluster and instead is formed in conjunction with the fused γ-lactam ring after generation and nucleophilic attack of a β-oxyanion[184]. The proposed formation of the β-lactone group in oxazolomycin biosynthesis involves intramolecular cyclization of a serine residue that is hypothesized to involve a terminal condensation domain instead of a thioesterase domain[183]. Ebelactone is a polyketide with a terminal β-lactone that is formed in a spontaneous manner without the use of a thioesterase[180]. Lipstatin also

contains a β-lactone group that is formed spontaneously following the reduction of a 3-keto group using a dehydrogenase[182]. In all of the examples mentioned above, β-lactone formation is one of the final steps in biosynthesis, and it is suggested that formation involves hydrolysis of the thioester that cleaves the natural product from the carrier protein. The variation in the enzymes proposed to be involved in the formation of these β-lactones suggests that a conserved method is not utilized by all bacteria for the formation of β-lactones in natural product biosynthesis. The formation of β-lactones in natural products remains unclear and is an area that requires further investigation to fully elucidate and characterize the mechanisms involved.

Given the noncanonical organization of the cystargolide gene cluster, the colinearity rule does not apply. This makes it difficult to predict the order that condensation reactions occur. It is possible that condensation occurs in a linear manner where the β-lactone precursor is the initiation subunit, which gets linked to valine, and successively to either isoleucine (cystargolide A) or a second valine (cystargolide B). It is also possible that the dipeptide is formed first and then gets linked to the β-lactone precursor subunit afterward. At this stage it is not clear what order the condensation reactions occur to produce cystargolides A and B (Figure 6.8), and thus both of these proposed mechanisms remain a possibility.

Given the evidence provided by the deduced biosyntheses of the β-lactone containing natural products ebelactone, salinasporamide A, oxazolomycin and lipstatin, β-lactone formation of cystargolide likely occurs via intramolecular condensation of a β-oxyanion and an activated carbonyl. The mechanisms involved in other β-lactone biosyntheses are often the final biosynthetic step, and involve cleavage of a thioester

bond that links the molecule to the carrier protein. The cystargolide gene cluster lacks a TE domain, and thus if this is the mechanism that is involved, cyclization and release would occur in a spontaneous manner like ebelactone and lipstatin or would be catalyzed by a different enzyme like oxazolomycin[180,182]. Cluster 34 contains a stand alone condensation domain that may catalyze this cyclization reaction. If formation of the β-lactone ring is caused by cleavage from the carrier protein, then two of the three adenylation domains present in the gene cluster would be responsible for activating both the 1-carboxyl group for condensation with the dipeptide, and the 4-carboxy group for intramolecular condensation and cyclization (Figure 6.9). This proposed mechanism would require that a single, promiscuous adenylation domain is responsible for activation of both isoleucine and valine for cystargolide A, and both valine residues for cystargolide B in an iterative manner. There are some genes in close proximity to those that encode the NRPSs in cluster 34 that suggest the involvement of a carrier-protein bound intermediate. ORF7564 encodes a gene with amino acid sequence homology to a 3-oxoacyl-ACP dehydratase/reductase (short chain dehydrogenase) that may be involved in modification of a CP-bound precursor (Figure 6.9). This family of enzymes is responsible for reduction of β-keto groups in fatty acid biosynthesis, and if it is involved in biosynthesis of cystargolide, could be responsible for reducing 2-isopropyl-3-oxosuccinate-ACP to 3-isopropyl malate-ACP. If this step is involved, then reduction of the β-keto followed by cyclization with the β-hydroxyl group would occur in a manner analogous to β-lactone formation in lipstatin. 2-Isopropyl-3-oxosuccinate is an intermediate in the leucine biosynthesis pathway and therefore could be recruited from primary metabolism for the production of cystargolide[185]. It may also be possible that

reduction of the β-keto group and cyclization occurs after condensation with the amino acids, since this family of enzymes is able to act on a variety of different substrates.

Alternatively, one adenylation domain could be responsible for activation of the 1-carboxyl group involved in condensation with the dipeptide chain, and the 4-carboxy group could be activated by a group other than an adenylate group. This proposed mechanism necessitates the involvement of additional modification enzymes for β-lactone formation. There are some genes in close proximity to those that encode the NRPSs in cluster 34 that may be involved in the modification of a cystargolide precursor or intermediate. The presence of a gene that encodes a putative acyl-CoA dehydrogenase (ORF7571) homolog located between the NRPS (ORF7572) and NRPS/PKS hybrid (ORF7570) suggests that formation of an acyl-CoA intermediate may be involved in biosynthesis. This is also supported by the presence of a gene encoding a putative enoyl-CoA hydratase (ORF7560) homolog that is located downstream of the gene cluster. If these genes are involved, the β-lactone could be formed by unsaturation of isopropyl succinate-CoA using acyl-CoA hydrogenase, followed by hydroxylation using enoyl-CoA hydratase to give 3-isopropyl malate-CoA. Cyclization could then occur through nucleophilic attack of the β-hydroxyl and cleavage of CoA to afford the final β-lactone containing subunit (Figure 6.9). Cluster 34 lacks an acyl-CoA ligase and thus if these enzymes are involved, an enzyme encoded elsewhere would have to be utilized to generate the CoA linked intermediate. This method involves the unusual isopropyl succinate as the initiation unit.

It may also be possible that none of these modification enzymes are required for the biosynthesis of cystargolide A and B. It still remains a possibility that cluster 34 is

only responsible for encoding the enzymes involved in the final stages of biosynthesis where the three subunits are linked together by NRPS domains, and the other enzymes involved are encoded at a different location in the genome. At this stage, it is unclear what the mechanism of cystargolide biosynthesis entails and what the boundaries are for the biosynthetic gene cluster. The mechanism of β-lactone formation and biosynthesis cannot be deduced from this gene cluster alone. Additional experiments including precursor feeding experiments, single gene knock outs and enzyme function assays would need to be performed to determine the mechanism of cystargolide biosynthesis.

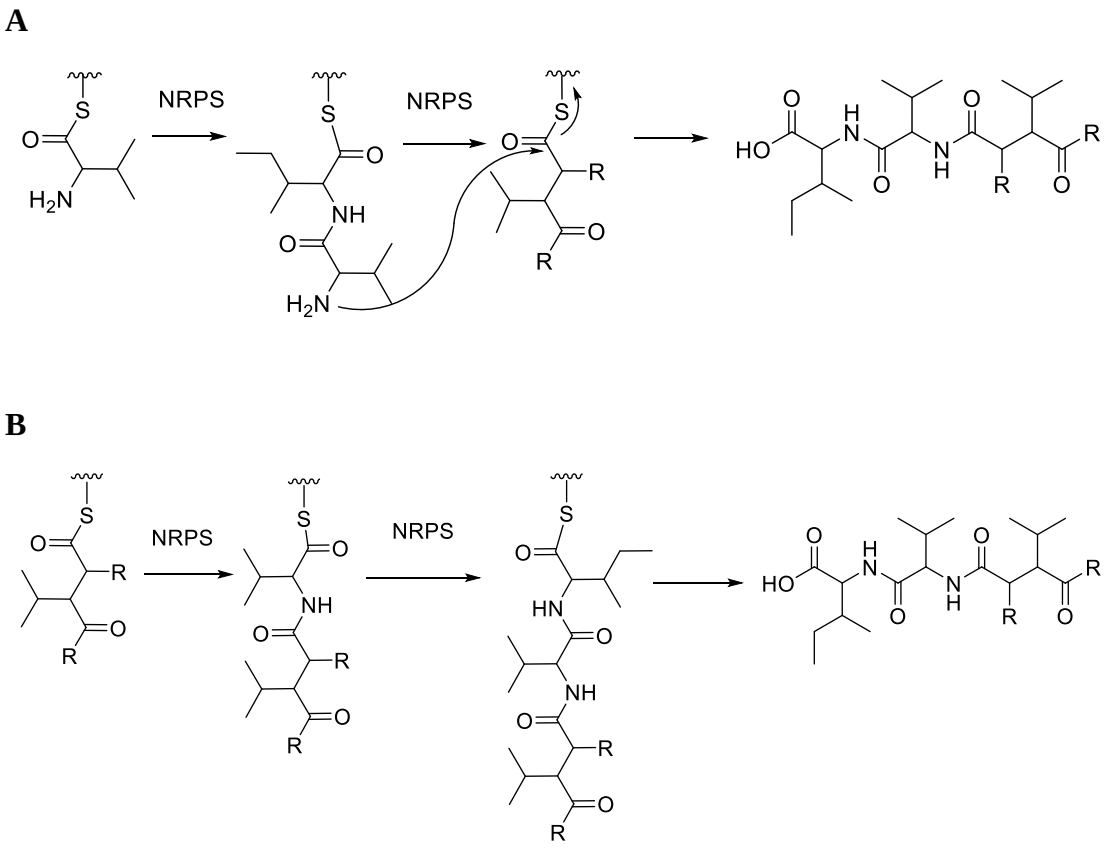
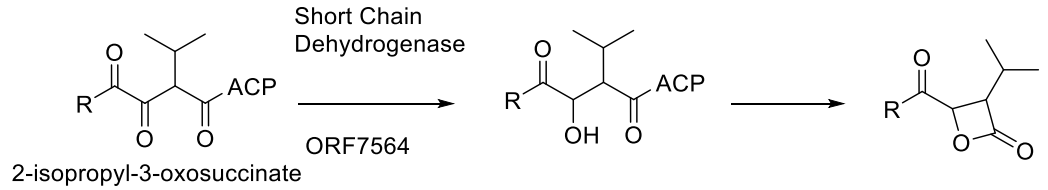


Figure 6.8. Possible methods of condensation between subunits for the biosynthesis of cystargolide.
A. A biosynthetic route that involves formation of the dipeptide first, followed by condensation with the β-lactone subunit precursor. R groups represent the unknown structures present at position 2 and position 5 during condensation. **B.** A biosynthetic route that involves the β-lactone subunit precursor as the initiation subunit, followed by condensdation with each successive amino acid in the dipeptide chain.

A



B

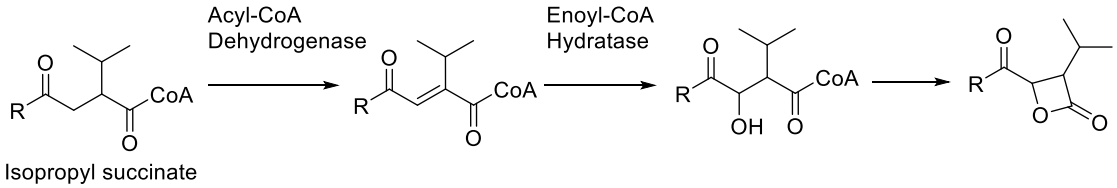


Figure 6.9. Possible intermediates involved in the biosynthesis of cystargolide. R groups could be –OH if the modifications occur before condensation or the dipeptide chain if they occur after condensation. **A.** Proposed mechanism if short chain dehydrogenase is involved. **B.** Proposed mechanism if acyl-CoA dehydrogenase and enoyl-CoA hydratase are involved.

6.3.4 Gene disruption mutants of the putative cystargolide gene cluster

Given the unique gene organization of cluster 34, and lack of co-linearity typically observed in the organization of NRPS genes, it was determined that bioinformatics alone would not be sufficient to determine if cluster 34 is involved in the biosynthesis of cystargolide A and B, and that gene disruption mutants would need to be generated for confirmation.

A *K. cystarginea* gene disruption mutant was created using the recently developed CRSIPR/cas9 genome editing technique[162]. The plasmid pCRISPomces-2 was designed to be efficiently altered to edit specific regions within the genome and was optimized for use in actinobacteria (Figure 6.10). In order to create a mutant strain that contained a small deletion in cluster 34, a sgRNA sequence that was adjacent to a PAM (NGG) sequence was selected as the protospacer region, and cloned into pCRISPomyces-2 to produce pCRISPCyOPS (CyOG2_PS_F and R). Two homologous template repair arms that were 0.85 kb upstream and downstream of the protospacer were amplified using PCR (CyOG2_L and R primer pairs) and then cloned into linear pCRISPCyOPS using the Gibson Assembly Method. The assembled plasmid was then introduced into *K. cystarginea* using conjugal DNA transfer from *E. coli* ET12567::pUZ8002, and selected for based on apramycin resistance. A total of 22 exconjugants were evaluated for the production of cystargolides A and B.

All exconjugants and the wild type strain were fermented using the cystargolide production conditions, and extracted with EtOAc. Extracts were analyzed by LC-MS for the presence or absence of cystargolide A and B. A single exconjugant (Ex-CyO12) produced no cystargolide A and B, and still retained production of cystargamide. To

ensure this positive result was reproducible, a spore stock of *K. cystarginea* Ex-CyO12 was prepared, and twelve replicates of the fermentations were repeated using single colonies from plates that were inoculated with the spore stock to ensure only a pure mutant culture. The extracts from the twelve replicate fermentations of *K. cystarginea* Ex-CyO12 were analyzed by LC-MS, which revealed production of cystargamide was retained, and production of cystargolide A and B was lost in all twelve replicates. This shows that the cystargolide biosynthetic gene cluster was disrupted and confirms the involvement of the locus in cystargolides biosynthesis. (Figure 6.11).

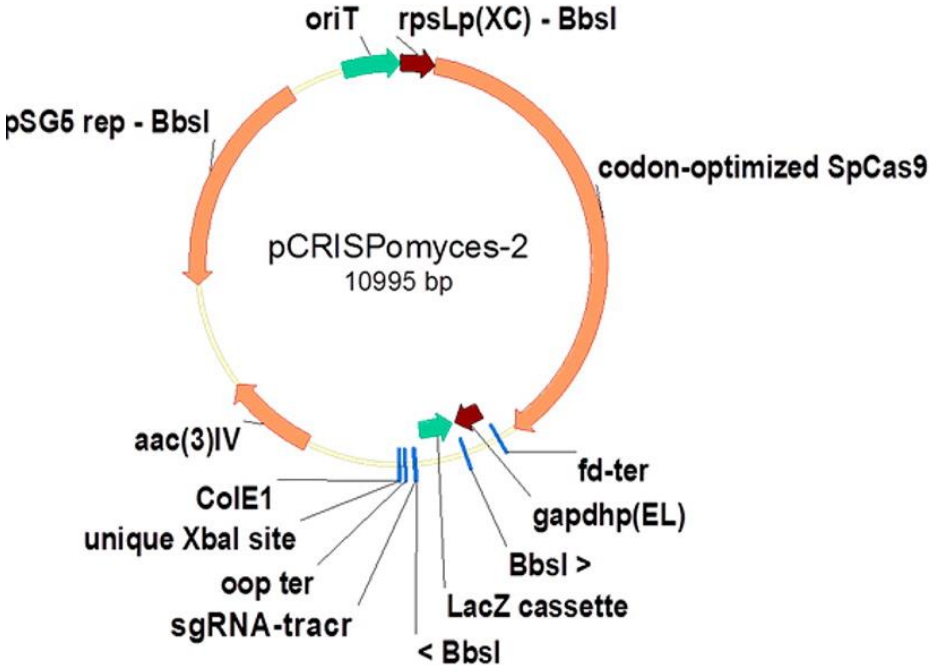


Figure 6.10. Plasmid map of pCRISPomyces-2 that was developed by R.E. Cobb for targeted genome editing in *Streptomyces*[162].

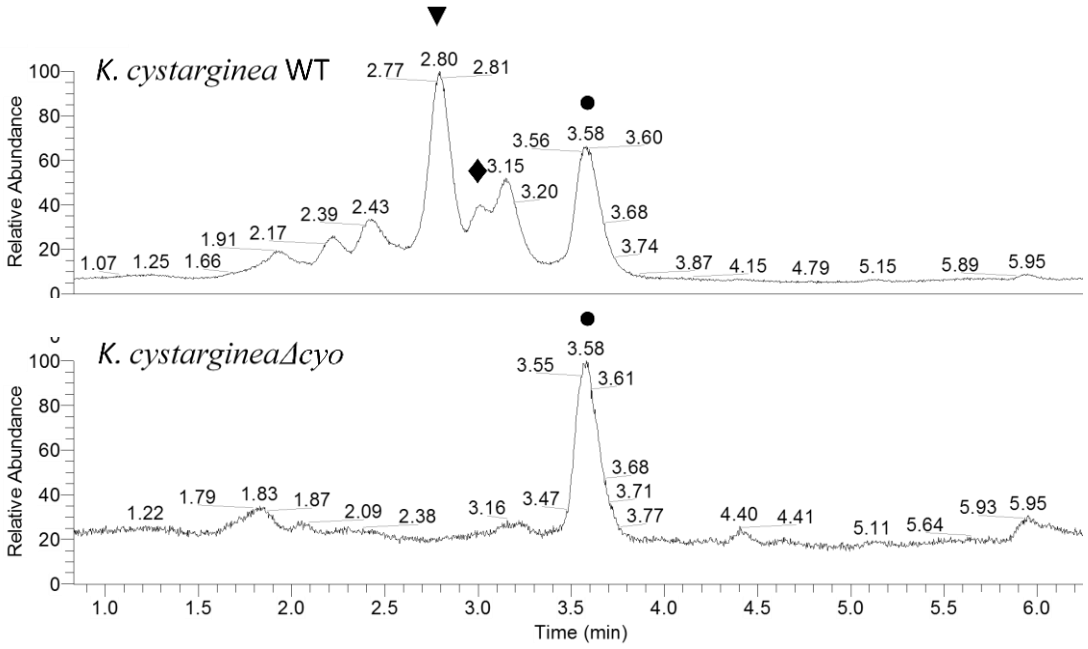


Figure 6.11. LCMS chromatograms of wild type *Kitasatospora cystarginea* and *K. cystargineaΔcyo*.
Retention of cystargamide (●) production in both strains, but loss of cystargolide A (♦) and B (▼) production compared to the wild type strain.

To confirm that the mutation in the genome of *K. cystarginea* ExCyO12 occurred in cluster 34, PCR was used to amplify a 1.2 kb region centered on the protospacer region (CyOG2_1.2_SF and R). The PCR primers used successfully amplified a 1.2 kb amplicon in the wild type strain, however were unable to amplify a smaller band expected from any of the replicates of the mutant strain. PCR was attempted again using a second set of primers to amplify a 200 bp amplicon (CyOG2_0.2_SF and R), as well as the homologous template primers (1.7 kb amplicon using CyOG2_FL and CyOG2_RR), and again both were unsuccessful at amplifying the mutant strain. These results suggested that the entire region of the genome that contains the homologous arms has been removed in the mutant strain instead of just the desired protospacer region. It is unlikely that the observed lack of cystargolide production was caused by contamination of another bacterium because cystargamide was still detected in all of the extracts. However, to ensure that contamination did not occur, the 16S rRNA sequences were amplified (16S_1525R and 16S_27F), sequenced (16S_936R) and compared to the wild type strain. BLASTn analysis revealed that all replicates of the *K. cystarginea* ExCyO12 and the wild type strain were *K. cystarginea.* This confirms that the lack of cystargolides production was caused by a gene mutation, and not contamination.

The unexpected results from PCR amplification suggested that the desired deletion of the protospacer region (20 bp) using the cas9 generated double strand break and homologous template repair did not occur in the anticipated manner to produce the mutant strains. Instead, it suggested that the gene mutation was much larger, and a region containing the entire 1.7 kb homologous template was altered. Instead of the expected method of homology directed repair (HDR), non-homologous end joining (NHEJ) could

have occurred, which could have introduced a random sequence where the double strand break occurred[162]. This method can introduce additional base pairs, but they are usually small additions. The large mutation could have also been a result of homologous recombination. At this time it is unclear what the size of the mutation is, but it is clear that the mutation occurred within cluster 34. Despite the unexpected method of mutation that occurred, cluster 34 was still rendered inactive. This resulted in loss of cystargolides production, revealing that cluster 34 is involved in the biosynthesis of cystargolide A and B.

6.4 CONCLUSIONS

The genome sequence of *K. cystaringea* revealed the presence of several putative natural product biosynthetic gene clusters. The clusters responsible for the production of the novel lipopeptide cystargamide were identified using bioinformatic annotations of putative gene functions. Cluster 11 displayed similarities to the calcium dependent antibiotic gene cluster, especially with regards to the genes involved in 2,3-epoxydecanoic acid biosynthesis and lipoinitiation. This cluster also contained the gene cassette required to produce the non-proteinogenic amino acid 4-hydroxyphenylglycine. Interestingly, the structure of cystargamide also contains the rare nonproteinogenic amino acid 5-hydroxytryptophan. The genes involved in hydroxylation of tryptophan to produce Htrp in bacteria are less extensively studied, and it is unclear if the gene is present within cluster 11.

As discussed in Chapter 4, cystargamide has demonstrated the ability to affect the metabolite production of other actinomycetes. Determining the gene cluster responsible for biosynthesis could allow for the generation of a library of analogs using synthetic

biology. The gene cluster could be altered so that different amino acids or fatty acids get incorporated. This could improve the biological activity and result in the generation of a more potent inducer molecule that could be used in drug discovery platforms to assist in the isolation of novel cryptic natural products from other actinomycetes.

The genome sequence of *K. cystarginea* was also used to determine the biosynthetic gene cluster responsible for producing the β-lactone containing natural products, cystargolides A and B. Given the unique structure of the cystargolides, bioinformatics alone was insufficient to determine the cystargolide gene cluster. The new CRISPR/cas9 genome editing technique was used to introduce a mutation to the genome within the predicted cystargolide gene cluster. Despite the unexpected size of the mutation that occurred, cluster 34 was still altered, which resulted in a loss of cystargolide A and B production. This shows that this gene cluster encodes enzymes that are involved in the biosynthesis of these natural products. Further investigation will be required to determine the method by which this gene disruption occurred. Not surprisingly, these unique natural products are produced by a unique gene cluster that differs from the typical genes that encode NRPs. The NRPS domains are encoded in a non-modular way, and do not follow the co-linearity rules of typical NRPS gene clusters. This makes it difficult to predict the order that condensation reactions between subunits occur, and makes it challenging to predict the mechanism involved in coupling of the peptide chain to the β-lactone subunit. The amino acids or subunits that are activated by two of the three adenylation domains could not be confidently predicted using bioinformatics software. This shows that they differ from the conserved regions of predictable adenylation domains and shows that they may display substrate promiscuity

or activate unusual subunits. This makes it challenging to predict if the β-lactone containing subunit is cyclized before condensation with the dipeptide occurs, or if it is modified and cyclized after. There are genes present in cluster 34 that may play a role in the biosynthesis of the substituted β-lactone subunit. A short chain dehydrogenase may be involved in reduction of a β-keto group to a β-hydroxyl that then cyclizes with a terminal carbonyl to cleave the carrier protein linked thioester. It could also be possible that an acyl-CoA dehydrogenase and enoyl-CoA hydratase are involved in β-unsaturation and hydroxylation, which then forms the β-lactone in a similar manner described above with cleavage of CoA. In order to determine the mechanism of biosynthesis and the boundaries of the cystargolide gene cluster, in frame single gene knockouts need to be generated, and the resulting intermediate metabolites would need to be characterized. Identification of the involvement of cluster 34 in the biosynthesis of cystargolides A and B is an exciting discovery that highlights the diverse biosynthetic ability of *Kitasatospora* bacteria. Determining the biosynthetic gene clusters of unique natural products helps to broaden our understanding of these capabilities. With the growing access and affordability of whole genome sequencing, genome based drug discovery programs are becoming more effective and more popular. The more information we know about BGCs, the more effective these genome based discovery programs will be. The divergence of the cystargolide gene cluster from the extensively characterized NRPS gene clusters reveals that there are still many natural product biosynthetic capabilities that remain unknown.

CHAPTER 7: CONCLUDING REMARKS AND FUTURE DIRECTIONS

7.1 Use of an LC-MS/PCA based screening approach for prioritization of bacterial libraries

In order to improve the efficiency of natural products drug discovery programs and avoid re-isolating known metabolites, new screening approaches that involve untargeted metabolomics have been recently developed[21,22]. By using high resolution LC-MS to analyze samples, and multivariate analyses like PCA to visualize the resulting data, strains that produce similar metabolites were grouped together, and strains that produce unique metabolites were prioritized. This method also provided information about the molecular weights of the metabolites, which allowed for the dereplication against known natural products, and assisted with the identification of putatively novel compounds. When this technique was applied to two bacterial libraries, it was successful at identifying two strains that produced three novel metabolites. Compared to traditional bioassay screening methods that were initially used, this technique was much more effective at identifying new natural products. Metabolites that were prioritized using bioassay results alone only resulted in the isolation of known compounds. These results show that using a chemo-centric approach to natural products discovery provided more information about the putative novelty of metabolites and allowed for better informed prioritization. This reduced the amount of time and resources spent reisolating known compounds. The metabolomics method is highly adaptable for high throughput platforms and thus can help improve the efficiency and success of natural product drug discovery programs.

This metabolomics based screening method was able to identify the production of three novel metabolites in two bacterial fermentation extracts. These metabolites are

among the first novel natural products to be identified and reported using this untargeted approach. Future expansions of this method are currently being investigated by other members of the Kerr lab to improve methods for the discovery of novel natural products. Recent advances have focused on adapting the multivariate analyses to include a targeted metabolomics approach that is used to highlight metabolites whose production is induced, or increased under different conditions. Several members of the Kerr lab are investigating methods to access cryptic natural products using techniques such as co-cultures, epigenetic modifiers and signalling molecules, and different stressors including salt, temperature variation or UV radiation. Continuing efforts are focusing on improving metabolomics based screening methods for identification of new compounds.

7.2 Characterization of novel natural products

One of the novel metabolites whose identification and characterization was described in this thesis was a lasso peptide natural product. This group of ribosomally synthesized post-translationally modified peptide natural products (RiPPs) is a relatively new group, and has garnered a lot of excitement because of their extreme stability under physiological conditions[84]. The threaded nature of lasso peptides makes them much more stable to degradation than linear or even cyclic peptides. Degradation of peptides by protease enzymes is one of the limiting factors that prevents many peptides from becoming clinically used drugs. The threaded structures of lasso peptides and the stability that these conformations afford helps to overcome this limitation and shows that they are exciting metabolites to investigate for their potential pharmaceutical applications.

The discovery of RKKD790_1592 represented a significant contribution to the group of lasso peptides because it was the smallest lasso peptide to be described. Its

structure provided information about the minimum requirements to create this 3D conformation. During the preparation of the manuscript describing the isolation of this interesting new metabolite, another research group published the isolation of this compound in the Journal of Natural Products[107], which was the recipient of the Jack L. Beal Award for excellence. This shows that the discovery of this new lasso peptide was an exciting discovery, and a significant contribution the field of natural products.

Another new metabolite that was identified and characterized was the lipopeptide, cystargamide, which was produced by the bacterial strain *K. cystarginea*. This natural product had interesting structural features including the nonproteinogenic amino acid 4-hydroxyphenyl glycine, the very rare amino acid 5-hydroxytryptophan and the interesting oxidized fatty acid, 2,3-epoxydecanoic acid. Preliminary results from our laboratory show that this metabolite displayed interesting effects on natural product production when added to fermentations of other actinomycetes. This demonstrates that cystargamide may be involved in bacterial communication. Cystargamide may be utilized as an inducer of secondary metabolite production, and could be applied to fermentations of large bacterial libraries in order to help access otherwise silent natural products. Future investigations of this exciting application will involve further characterization of the induction properties including determination of the molecular target and cellular pathway that cystargamide interacts with, and application to a larger number of actinomycetes and other bacterial phyla.

Cystargolides A and B are β-lactone containing peptides that were also isolated from the fermentation broth of *K. cystarginea.* These novel metabolites contained a unique β-lactone group that bears isopropyl and carboxyl substituents. There are very few

natural products that have been previously described that have similar structural features to cystargolide A and B. The characterization of these very unique natural products greatly increases our knowledge of the structural diversity of natural products that bacteria are capable of biosynthesizing.

Cystargolide A and B displayed potent inhibition of the human 20S proteasome in an i*n vitro* enzyme assay. When evaluated in cancer cell-line based assays, the potency of these compounds decreased. Future studies involving a medicinal chemistry approach will aim at increasing the potency of these compounds and improving the *in vivo* anticancer activity. Modification of the amino acid side chains, as well as the alkyl substituent on the β-lactone ring are being evaluated for their potential to increase activity. The reactivity and lability of the β-lactone might be the cause of decreased potency in the cell line assays so a series of β-lactam analogs are also being produced to try and increase stability, but retain inhibition of the proteasome. These studies are being conducted by several members of the Kerr lab including an undergraduate student, a postdoctoral fellow, and a laboratory manager.

The discovery of these novel metabolites has attracted the attention of several other research groups. The group led by Dr. Tello-Aburto has completed a total synthesis of these β-lactone containing peptide[147]. In addition to the synthesized natural products, this research group also evaluated a benzyl ester derivative for antiproliferative activity, which resulted in IC_{50} values 100 times more potent that the natural products. This shows that simple derivatizations can greatly improve activity and thus there is potential for further structural modifications. The discovery of these natural products has also led to a collaboration with Dr. Barrie Wilkinson who is conducting isotopically labelled precursor

feeding experiments to determine biosynthetic intermediates, and mechanisms responsible for the biosynthesis of cystargolide A and B. This shows that there is interest in these compounds in the scientific community and that their discovery significantly contributes to the field of natural products.

7.3 Identification of biosynthetic gene clusters in the genome sequence of *Kitastospora cystarginea*

The biosynthetic gene clusters of cystargamide and cystargolides were identified from the genome sequence of *K. cystarginea* using bioinformatics predictions and gene disruption techniques. Identifying the biosynthetic gene clusters responsible for producing these novel natural products is important because it allows us to better understand the basic principles of how natural products are produced by bacteria. The more knowledge that is gained about the genes and mechanisms involved in the biosynthesis of metabolites, the better able we are to predict unknown gene clusters, and the metabolites they produce. This allows for advances in genome based natural products discovery programs, and contributes to the development of improved ways of natural products based drug discovery programs[18]. It also allows for the generation of analog libraries using synthetic biology techniques. This helps to create analogs that are unattainable using synthetic approaches, which may contribute to the development of analogs with improved potency or other pharmacological properties. Determining the biosynthetic gene clusters also allows for genetic engineering for strain optimization and improved yields. This is important for commercialization of strains used to produce mass quantities of a drug candidate. Determining the biosynthetic gene clusters responsible for producing cystargamide and cystargolides in the genome sequence of *K. cystarginea* has

increased the basic understanding of bacterial genetics, revealed insights into the mechanisms of biosynthesis of these novel metabolites, and opened the doors for future experiments to determine the boundaries of these gene clusters and enzyme function encoded by each gene present in the gene clusters. The Wilkinson research group is currently undertaking these labelled feeding experiments to ascertain subunits and intermediates involved in the biosynthesis of cystargolides A and B. Together with the gene cluster that was identified, the results of these experiments will reveal the biosynthesis of these unique natural products.

7.4 Summary

The research described in this thesis outlines the discovery of three novel bacterial natural products using an innovative LC-MS/PCA based screening approach. The structures of these natural products were elucidated and their potential biological applications were assessed in several bioassays. It was determined that the lipopeptide, cystargamide displayed promising activity as a bacterial signalling molecule that may be used to induce metabolite production in other bacteria, and the β-lactone containing peptides, cystargolides A and B are inhibitors of the 20S proteasome that may be used as antiproliferating agents. The whole genome sequence of the producing strain *K. cystarginea* was sequenced, annotated, and the gene cluster responsible for production of these novel metabolites were identified. This discovery of these novel metabolites from this unique strains shows that members of the genus *Kitasatospora* represent a prolific and underexplored resource for the discovery of novel compounds. These novel natural products and their biosynthetic gene clusters have inspired several other research projects in the Kerr lab, and other labs internationally. This shows that the discovery of these

metabolites, their characterization and determination of their biosynthesis is an exciting and significant contribution to the field of microbial natural products.

REFERENCES

1. Stone, M. J. & Williams, D. H. On the evolution of functional secondary metabolites (natural products). *Mol Microbiol* **6,** 29–34 (1992).

2. Osbourn, A. E. & Lanzotti, V. *Plant-derived Natural Products*. (Springer Science & Business Media, 2009).

3. Buckingham, J. *Dictionary of Natural Products*. (CRC Press, 1993).

4. McClintock, J. B. & Baker, B. J. *Marine Chemical Ecology*. (CRC Press, 2001).

5. Simmonds, M. S. J. Flavonoid-insect interactions: recent advances in our knowledge. *Phytochem.* **64,** 21–30 (2003).

6. Dixon, R. A. Natural products and plant disease resistance. *Nature* **411,** 843–847 (2001).

7.Olivera, B. M., Gray, W. R., Zeikus, R. & McIntosh, J. M. Peptide neurotoxins from fish-hunting cone snails. *Science* **230,** 1338–1343 (1985).

8. Williams, P. Quorum sensing, communication and cross-kingdom signalling in the bacterial world. *Microbiol* **153,** 3923–3938 (2007).

9. Bossier, P., Hofte, M. & Verstraete, W. *Advances in Microbial Ecology* (Marshall, K. C.) **10,** 385–414 (Springer US, 1988).

10. Qian, P. Y., Xu, Y. & Fusetani, N. Natural products as antifouling compounds: recent progress and future perspectives. *Biofouling* **26,** 223–234 (2010).

11. Newman, D. J. & Cragg, G. M. Natural Products as Sources of New Drugs from 1981 to 2014. *J Nat Prod* **79,** 629–661 (2016).

12. World Health Organization. *Antimicrobial resistance: global report on surveillance*. 1–257 (2014).

13. Patridge, E., Gareiss, P., Kinch, M. S. & Hoyer, D. An analysis of FDA-approved drugs: natural products and their derivatives. *Drug Disc Today* **21,** 204–207 (2016).

14. Bérdy, J. Bioactive microbial metabolites. *J Antibiot* **58,** 1–26 (2005).

15. Demain, A. L. Importance of microbial natural products and the need to revitalize their discovery. *J Ind Microbiol Biotechnol* **41,** 185–201 (2014).

16. Koehn, F. E. & Carter, G. T. The evolving role of natural products in drug discovery. *Nat Rev Drug Disc* **4,** 206–220 (2005).

17. Luo, Y., Enghiad, B. & Zhao, H. New tools for reconstruction and heterologous expression of natural product biosynthetic gene clusters. *Nat Prod Rep* **33,** 174–182 (2015).

18. Harvey, A. L., Edrada, R. A. & Quinn, R. J. The re-emergence of natural products for drug discovery in the genomics era. *Nat Rev Drug Disc* **14,** 111–129 (2015).

19. Katz, L. & Baltz, R. H. Natural product discovery: past, present, and future. *J Ind Microbiol Biotechnol* **43,** 155–176 (2016).

20. Kurita, K. L. & Linington, R. G. Connecting phenotype and chemotype: high-content discovery strategies for natural products research. *J Nat Prod* **78,** 587–596 (2015).

21.Forner, D., Berrué, F., Correa, H., Duncan, K. & Kerr, R. G. Chemical dereplication of marine actinomycetes by liquid chromatography–high resolution mass spectrometry profiling and statistical analysis. *Anal Chim Acta* **805,** 70–79 (2013).

22. Hou, Y. *et al.* Microbial Strain Prioritization Using Metabolomics Tools for the Discovery of Natural Products. *Anal Chem* **84,** 4277 (2012).

23. Robinette, S. L., Bruschweiler, R., Shroeder, F. C. & Edison, A. S. NMR in metabolomics and natural products research: two sides of the same coin. *Acc Chem Res* **45,** 288–297 (2011).

24. Sanchez, L. M. *et al.* Molecular networking as a dereplication strategy. *J Nat Prod* **76,** 1686–1699 (2013).

25. Lentzen, G. & Schwarz, T. Extremolytes: natural compounds from extremophiles for versatile applications. *Appl Microbiol Biotech* **72,** 623–634 (2006).

26. Molinski, T. F., Dalisay, D. S., Lievens, S. L. & Saludes, J. P. Drug development from marine natural products. *Nat Rev Drug Disc* **8,** 69–85 (2009).

27. Scherlach, K. & Hertweck, C. Triggering cryptic natural product biosynthesis in microorganisms. *Org Biomol Chem* **7,** 1753–1760 (2009).

28. Pettit, R. K. Mixed fermentation for natural product drug discovery. *Appl Microbiol Biotech* **83,** 19–25 (2009).

29. El-Elimat, T. *et al.* Flavonolignans from *Aspergillus iizukae*, a fungal endophyte of milk thistle (*Silybum marianum*). *J Nat Prod* **77,** 193–199 (2014).

30. Kinghorn, A. D. *et al.* Cancer chemopreventive agents discovered by activity-guided fractionation. *Curr Org Chem* **7,** 597–612 (1998).

31. Kellogg, J. J. *et al.* Biochemometrics for Natural Products Research: Comparison of Data Analysis Approaches and Application to Identification of Bioactive Compounds. *J Nat Prod* **79,** 376–386 (2016).

32. Kleigrewe, K. *et al.* Combining mass spectrometric metabolic profiling with genomic analysis: a powerful approach for discovering natural products from cyanobacteria. *J Nat Prod* **78,** 1671–1682 (2015).

33. Sidebottom, A. M., Johnson, A. R., Karty, J. A., Trader, D. J. & Carlson, E. E. Integrated metabolomics approach facilitates discovery of an unpredicted natural product suite from *Streptomyces coelicolor* M145. *ACS Chem Biol* **8,** 2009–2016 (2013).

34. Rajalahti, T. & Kvalheim, O. M. Multivariate data analysis in pharmaceutics: a tutorial review. *Int J Pharma* **417,** 280–290 (2011).

35. Demain, A. L. & Sanchez, S. Microbial drug discovery: 80 years of progress. *J Antibiot* **62,** 5–16 (2009).

36. Nelson, M. L. & Levy, S. B. The history of the tetracyclines. *Ann NY Acad Sci* **1241,** 17–32 (2011).

37. Levine, D. P. Vancomycin: a history. *Clin Infect Dis* **42,** 5–12 (2006).

38. Sauermann, R., Rothenburger, M. & Graninger, W. Daptomycin: a review 4 years after first approval. *Pharmacology* **81,** 79–91 (2008).

39. Bentley, S. D. *et al.* Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3 (2). *Nature* **417,** 141–147 (2002).

40. Omura, S., Takahashi, Y., Iwai, Y. & Tanaka, H. *Kitasatosporia*, a new genus of the order *Actinomycetales*. *J Antibiot* **35,** 1013–1019 (1982).

41. Ludwig, W. *et al.* in *Bergey's Manual of Systematic Bacteriology* 1–31 (2012).

42. Wellington, E., Stackebrandt, E., Sanders, D., Wolstrup, J. & Jordensen, N. O. G. Taxonomic status of *Kitasatosporia*, and proposed unification with *Streptomyces* on the basis of phenotypic and 16S rRNA analysis and emendation of *Streptomyces*. *Int J Sys Bacteriol* **42,** 156–160 (1992).

43. Zhang, Z., Wang, Y. & Ruan, J. A proposal to revive the genus *Kitasatospora* (Omura, Takahashi, Iwai, and Tanaka 1982). *Int J Syst Bacteriol* **47,** 1048–1054 (1997).

44. Hwang, J. Y., Kim, S. H., Oh, H. R., Kwon, E. & Nam, D. H. Analysis of a draft genome sequence of *Kitasatospora cheerisanensis* KCTC 2395 producing bafilomycin antibiotics. *J Microbiol* **53,** 84–89 (2014).

45. Girard, G. *et al.* Analysis of novel *kitasatosporae* reveals significant evolutionary changes in conserved developmental genes between *Kitasatospora* and *Streptomyces*. *Antonie Van Leeuwenhoek* **106,** 365–380 (2014).

46. Ichikawa, N. *et al.* Genome Sequence of *Kitasatospora setae* NBRC 14216T: An Evolutionary Snapshot of the Family *Streptomycetaceae*. *DNA Res* **17,** 393–406 (2010).

47. Duncan, K., Haltli, B., Gill, K. A. & Kerr, R. G. Bioprospecting from marine sediments of New Brunswick, Canada: exploring the relationship between total bacterial diversity and actinobacteria diversity. *Mar Drugs* **12,** 899–925 (2014).

48. Duncan, K. R. *et al.* Exploring the diversity and metabolic potential of actinomycetes from temperate marine sediments from Newfoundland, Canada. *J Ind Microbiol Biotech* **42,** 57–72 (2015).

49. Shirling, E. B. & Gottlieb, D. Methods for characterization of *Streptomyces* species. *Int J Syst Bacteriol* **16,** 313–340 (1966).

50. Gill, K. A., Berrué, F., Arens, J. C. & Kerr, R. G. Isolation and structure elucidation of cystargamide, a lipopeptide from *Kitasatospora cystarginea*. *J Nat Prod* **77,** 1372–1376 (2014).

51. Gill, K. A., Berrué, F., Arens, J. C., Carr, G. & Kerr, R. G. Cystargolides, 20S Proteasome Inhibitors Isolated from *Kitasatospora cystarginea*. *J Nat Prod* **78,** 822–826 (2015).

52. Pluskal, T., Castillo, S., Villar-Briones, A. & Oresic, M. MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* **11,** 395–406 (2010).

53. Methods for dilution antimicrobial susceptibility-Test for bacteria that grow aerobically. *National Committee for Clinical Laboratory Standards* **CLSI document M07-A9,** Wayne, PA–9th Edition (2012).

54. Altuschul, S., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Blast (basic local alignment search tool). *JMol Biol* **215,** 403–410

55. Stach, J. *et al.* Statistical approaches for estimating actinobacterial diversity in marine sediments. *Appl Environ Microbiol* **69,** 6189–6200 (2003).

56. Decker, H. *et al.* A general approach for cloning and characterizing dNDP-glucose dehydratase genes from actinomycetes. *FEMS Microbiol Lett* **141,** 195–201 (1996).

57. Goldstein, J. L. & Brown, M. S. Regulation of the mevalonate pathway. *Nature* **343,** 425–430 (1990).

58. Hopwood, D. A. Complex Enzymes in Microbial Natural Product Biosynthesis. *Methods in Enzymology* (Elsevier, 2009).

59.Miyairi, N., Sakai, H., Konomi, T. & Imanaka, H. Enterocin, a new antibiotic taxonomy, isolation and characterization. *J Antibiot* **29,** 227–235 (1976).

60. Mohimani, H. *et al.* Multiplex de novo sequencing of peptide antibiotics. *J Comp Biol* **18,** 1371–1381 (2011).

61. Furusaki, A. *et al.* X-ray crystal structure of staurosporine: a new alkaloid from a *Streptomyces* strain. *J Chem Soc Chem Commun* **18,** 800–801 (1978).

62. Omura, S. *et al.* A new alkaloid AM-2282 of *Streptomyces* origin taxonomy, fermentation, isolation and preliminary characterization. *J Antibiot* **30,** 275–282 (1977).

63. Nakano, H. Chemical biology of natural indolocarbazole products: 30 years since the discovery of staurosporine. *J Antibiot* **62,** 17–26 (2009).

64. Hoeksema, H., Johnson, J. L. & Hinman, J. W. Structural studies on streptonivicin, 1 a new antibiotic. *J Am Chem Soc* **77,** 6710–6711 (1955).

65. Smith, C. G., Dietz, A., Sokolski, W. T. & Savage, G. M. Streptonivicin, a new antibiotic. I. Discovery and biologic studies. *Antibiot & Chemother* **6,** 135–142 (1956).

66. Iakobson, L. M. The use of antibiotics; novobiocin (streptonivicin, catomycin, catocin, cardelmycin, albamycin. *Antibiotiki* **10,** 46–52 (1957).

67. Dobler, M. The crystal structure of nonactin. *Helv Chim Acta* **55,** 1371–1384 (1972).

68. Graven, S. N., Lardy, H. A., Johnson, D. & Rutter, A. Antibiotics As Tools for Metabolic Studies. V. Effect of Nonactin, Monactin, Dinactin, and Trinactin on Oxidative Phosphorylation and Adenosine Triphosphatase Induction. *Biochemistry* **5,** 1729–1735 (1966).

69. Schimana, J. *et al.* Arylomycins A and B, New Biaryl-bridged Lipopeptide Antibiotics Produced by *Streptomyces* sp. Tue 6075. I. Taxonomy, Fermentation, Isolation and Biological Activities. *J Antibiot* **55,** 565–570 (2002).

70. Schleissner, C. *et al.* Antitumor actinopyranones produced by *Streptomyces albus* POR-04-15-053 isolated from a marine sediment. *J Nat Prod* **74,** 1590–1596 (2011).

71. Yano, K. *et al.* Actinopyrones A, B and C, new physiologically active substances. I. Producing organism, fermentation, isolation and biological properties. *J Antibiot* **39,** 32–37 (1986).

72. Yano, K. *et al.* Actinopyrones A, B and C, new physiologically active substances. II. Physico-chemical properties and chemical structures. *J Antibiot* **39,** 38–43 (1986).

73. Pinkerton, M., Steinrauf, L. K. & Dawkins, P. The molecular structure and some transport properties of valinomycin. *Biochem Biophys Res Comm* **34,** 512–518 (1969).

74. Neupert-Laves, K. & Dobler, M. The crystal structure of a K+ complex of valinomycin. *Helv Chim Acta* **58,** 432–442 (1975).

75. Shemyakin, M. M., Vinogradova, E. I. & Feigina, M. Y. The structure-antimicrobial relation for valinomycin depsipeptides. *Experientia* **21,** 548–552 (1965).

76. Arens, J. C., Berrué, F., Pearson, J. K. & Kerr, R. G. Isolation and Structure Elucidation of Satosporin A and B: New Polyketides from *Kitasatospora griseola*. *Org Lett* **15,** 3864–3867 (2013).

77. Huang, H. *et al.* Antimalarial β-carboline and indolactam alkaloids from *Marinactinospora thermotolerans*, a deep sea isolate. *J Nat Prod* **74,** 2122–2127 (2011).

78. Woo, E. J., Starks, C. M., Carney, J. R. & Arslanian, R. Migrastatin and a new compound, isomigrastatin, from *Streptomyces platensis*. *J Antibiot* **55,** 141–146 (2002).

79. Habib, E., Yokomizo, K., Murata, K. & Uyeda, M. Structures of fattiviracin family, antiviral antibiotics. *J Antibiot* **53,** 1420–1423 (2000).

80. Williams, T. H. Naphthomycin, a novel ansa macrocyclic antimetabolite. Proton NMR spectra and structure elucidation using lanthanide shift reagent. *J Antibiot* **28,** 85–86 (1975).

81. Umezawa, H. *et al.* Pepstatin, a new pepsin inhibitor produced by actinomycetes. *J Antibiot* **23,** 259–262 (1970).

82. Yang, X. & van der Donk, W. A. Ribosomally Synthesized and Post-Translationally Modified Peptide Natural Products: New Insights into the Role of Leader and Core Peptides during Biosynthesis. *Chem Europ J* **19,** 7662–7677 (2013).

83. Arnison, P. G. *et al.* Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat Prod Rep* **30,** 108–160 (2012).

84. Maksimov, M. O., Pan, S. J. & Link, A. J. Lasso peptides: structure, function, biosynthesis, and engineering. *Nat Prod Rep* **29,** 996–1006 (2012).

85. Hegemann, J. D., Zimmermann, M., Xie, X. & Marahiel, M. A. Lasso peptides: an intriguing class of bacterial natural products. *Acc Chem Res* **48,** 1909–1919 (2015).

86. Fréchet, D. *et al.* Solution structure of RP 71955, a new 21 amino acid tricyclic peptide active against HIV-1 virus. *Biochemistry* **33,** 42–50 (1994).

87. Katahira, R., Shibata, K., Yamasaki, M., Matsuda, Y. & Yoshida, M. Solution structure of endothelin B receptor selective antagonist RES-701-1 determined by 1H NMR spectroscopy. *Bioorg Med Chem* **3,** 1273–1280 (1995).

88. Constantine, K. L. *et al.* High-resolution solution structure of siamycin II: novel amphipathic character of a 21-residue peptide that inhibits HIV fusion. *J Biomol NMR* **5,** 271–286 (1995).

89. Bayro, M. J. *et al.* Structure of antibacterial peptide microcin J25: a 21-residue lariat protoknot. *J Am Chem Soc* **125,** 12382–12383 (2003).

90. Rosengren, K. J. *et al.* Microcin J25 has a threaded sidechain-to-backbone ring structure and not a head-to-tail cyclized backbone. *JAm Chem Soc* **125,** 12464–12474 (2003).

91. Iwatsuki, M. *et al.* Lariatins, antimycobacterial peptides produced by *Rhodococcus* sp. K01-B0171, have a lasso structure. *J Am Chem Soc* **128,** 7486–7491 (2006).

92. Knappe, T. A. *et al.* Isolation and structural characterization of capistruin, a lasso peptide predicted from the genome sequence of *Burkholderia thailandensis* E264. *J Am Chem Soc* **130,** 11446–11454 (2008).

93. Hegemann, J. D., Zimmermann, M., Xie, X. & Marahiel, M. A. Caulosegnins I-III: a highly diverse group of lasso peptides derived from a single biosynthetic gene cluster. *J Am Chem Soc* **135,** 210–222 (2012).

94. Knappe, T. A., Linne, U., Xie, X. & Marahiel, M. A. The glucagon receptor antagonist BI-32169 constitutes a new class of lasso peptides. *FEBS Lett* **584,** 785–789 (2009).

95. Potterat, O. *et al.* BI-32169, a bicyclic 19-peptide with strong glucagon receptor antagonist activity from *Streptomyces* sp. *J Nat Prod* **67,** 1528–1531 (2004).

96. Soudy, R., Wang, L. & Kaur, K. Synthetic peptides derived from the sequence of a lasso peptide microcin J25 show antibacterial activity. *Bioorg Med Chem* **20,** 1794–1800 (2012).

97. Sato, A. K., Viswanathan, M., Kent, R. B. & Wood, C. R. Therapeutic peptides: technological advances driving peptides into development. *Curr Opin Biotechnol* **17,** 638–642 (2006).

98. Antosova, Z., Mackova, M., Kral, V. & Macek, T. Therapeutic application of peptides and proteins: parenteral forever? *Trends Biotech* **27,** 628–635 (2009).

99. Sasse, F., Kessler, H., Xie, X. & Marahiel, M. A. Introducing lasso peptides as molecular scaffolds for drug design: engineering of an integrin antagonist. *Angew Chem Int Ed* **50,** 8714–8717 (2011).

100. Marfey, P. Determination of D-amino acids. II. Use of a bifunctional reagent, 1, 5-difluoro-2, 4-dinitrobenzene. *Carl Res Comm* **49,** 591 (1984).

101. Stewart, J. J. P. Optimization of parameters for semiempirical methods I. Method. *J Comp Chem* **10,** 209–220 (1989).

102. Stewart, J. J. P. Optimization of parameters for semiempirical methods II. Applications. *J Comp Chem* **10,** 221–264 (1989).

103. Bauer, A. W., Kirby, W., Sherris, J. C. & Turck, M. Antibiotic susceptibility testing by a standardized single disk method. *Am J Clin Pathol* **45,** 493–496 (1966).

104. Guo, Y., Zheng, W., Rong, X. & Huang, Y. A multilocus phylogeny of the *Streptomyces griseus* 16S rRNA gene clade: use of multilocus sequence analysis for *Streptomycete* systematics. *Int J Syst Evol Bacteriol* **58,** 149–159 (2008).

105. Johnson, R. S., Martin, S. A., Biemann, K., Stults, J. T. & Watson, J. T. Novel fragmentation process of peptides by collision-induced decomposition in a tandem mass spectrometer: differentiation of leucine and isoleucine. *Anal Chem* **59,** 2621–2625 (1987).

106. Weber, W., Fischli, W., Hochuli, E., Kupfer, E. & Weibel, E. K. Anantin--a peptide antagonist of the atrial natriuretic factor (ANF). I. Producing organism, fermentation, isolation and biological activity. *J Antibiot* **44,** 164–171 (1991).

107. Um, S. *et al.* Sungsanpin, a lasso peptide from a deep-sea *Streptomycete*. *J Nat Prod* **76,** 873–879 (2013).

108. Mootz, H. D., Schwarzer, D. & Marahiel, M. A. Ways of assembling complex natural products on modular nonribosomal peptide synthetases. *Chembiochem* **3,** 490–504 (2002).

109. Cane, D. E., Walsh, C. T. & Khosla, C. Harnessing the biosynthetic code: combinations, permutations, and mutations. *Science* **282,** 63–68 (1998).

110. Marahiel, M. A., Stachelhaus, T. & Mootz, H. D. Modular Peptide Synthetases Involved in Nonribosomal Peptide Synthesis. *Chem Rev* **97,** 2651–2674 (1997).

111. Baltz, R. H., Miao, V. & Wrigley, S. K. Natural products to drugs: daptomycin and related lipopeptide antibiotics. *Nat Prod Rep* **22,** 717–741 (2005).

112. Strieker, M. & Marahiel, M. A. The structural diversity of acidic lipopeptide antibiotics. *Chembiochem* **10,** 607–616 (2009).

113. Debono, M. *et al.* A21978C, a complex of new acidic peptide antibiotics. Isolation, chemistry, and mass spectral structure elucidation. *J Antibiot* **40,** 761–777 (1987).

114. Hopwood, D. A. & Wright, H. M. CDA is a new chromosomally-determined antibiotic from *Streptomyces coelicolor* A3 (2). *Microbiology* **129,** 3575–3579 (1983).

115. Aretz, W., Meiwes, J., Seibert, G., Vobis, G. & Wink, J. Friulimicins: novel lipopeptide antibiotics with peptidoglycan synthesis inhibiting activity from *Actinoplanes friuliensis* sp. nov. I. Taxonomic studies of the producing microorganism and fermentation. *J Antibiot* **53,** 807–815 (2000).

116. Naganawa, H. *et al.* Laspartomycin, a new anti-*Staphylococcal* peptide. *J Antibiot* **21,** 55–62 (1968).

117. Kakinuma, A., Sugino, H., Isono, M., Tamura, G. & Arima, K. Determination of fatty acid in surfactin and elucidation of the total structure of surfactin. *Ag Biol Chem* **33,** 973–976 (1969).

118. Zahner, H. *et al.* Metabolic products of microorganisms. 224. Bafilomycins, a new group of macrolide antibiotics. Production, isolation, chemical structure and biological activity. *J Antibiot* **37,** 110–117 (1984).

119. Yoon, T. M., Kim, J. W., Kim, J. G., Kim, W. G. & Suh, J. W. Talosins A and B: new isoflavonol glycosides with potent antifungal activity from *Kitasatospora kifunensis* MJM341. I. Taxonomy, fermentation, isolation, and biological activities. *J Antibiot* **59,** 633–639 (2006).

120. Hiramoto, M. *et al.* Structures of Novel Phospholipase A2 Inhibitors Q-6402 A and B. *Symposium of the Chemistry of Natural Products* 678–684 (1993).

121.Castiglione, F., Marazzi, A., Meli, M. & Colombo, G. Structure elucidation and 3D solution conformation of the antibiotic enduracidin determined by NMR spectroscopy and molecular dynamics. *Magn Reson Chem* **43,** 603–610 (2005).

122. Somner, E. A. & Reynolds, P. E. Inhibition of peptidoglycan biosynthesis by ramoplanin. *Antimicrob Agents Chemother* **34,** 413–419 (1990).

123. Slominski, A. *et al.* Conversion of L-tryptophan to serotonin and melatonin in human melanoma cells. *FEBS Lett* **511,** 102–106 (2002).

124. Hojati, Z. *et al.* Structure, biosynthetic origin, and engineered biosynthesis of calcium-dependent antibiotics from *Streptomyces coelicolor*. *Chem Biol* **9,** 1175–1187 (2002).

125. Williamson, K. L., Lanford, C. A. & Nicholson, C. R. The Nuclear Magnetic Resonance Spectra of Three-Membered Ring Compounds: Substituent Effects on

Monosubstituted 1,1-Dichlorocyclopropanes and Epoxides. *J Am Chem Soc* **86,** 762–765 (1964).

126. Straight, P. D., Willey, J. M. & Kolter, R. Interactions between *Streptomyces coelicolor* and *Bacillus subtilis*: Role of surfactants in raising aerial structures. *J Bacteriol* **188,** 4918–4925 (2006).

127. Bibb, M. The regulation of antibiotic production in *Streptomyces coelicolor* A3 (2). *Microbiology* **142,** 1335–1344 (1996).

128. Weibel, E. K. *et al.* Lipstatin, an inhibitor of pancreatic lipase, produced by *Streptomyces toxytricini*. I. Producing organism, fermentation, isolation and biological activity. *J Antibiot* **40,** 1081–1085 (1987).

129. Guerciolini, R. Mode of action of orlistat. *Int J Obes Relat Metab Disord* **21,** 12–23 (1997).

130. Asai, A., Hasegawa, A., Ochiai, K., Yamashita, Y. & Mizukami, T. Belactosin A, a novel antitumor antibiotic acting on cyclin/CDK mediated cell cycle regulation, produced by *Streptomyces* sp. *J Antibiot* **53,** 81–83 (2000).

131. Asai, A. *et al.* A new structural class of proteasome inhibitors identified by microbial screening using yeast-based assay. *Biochem Pharmacol* **67,** 227–234 (2003).

132. Feling, R. H. *et al.* Salinosporamide A: a highly cytotoxic proteasome inhibitor from a novel microbial source, a marine bacterium of the new genus *Salinospora*. *Angew Chem Int Ed Engl* **42,** 355–357 (2003).

133. Umezawa, H. *et al.* Ebelactone, an inhibitor of esterase, produced by actinomycetes. *J Antibiot* **33,** 1594–1596 (1980).

134. Orlowski, M. The multicatalytic proteinase complex, a major extralysosomal proteolytic system. *Biochemistry* **29,** 10289–10297 (1990).

135. Fuchs, S. Y. The role of ubiquitin-proteasome pathway in oncogenic signaling. *Cancer Biol Ther* **1,** 337–341 (2002).

136. Muchamuel, T. *et al.* A selective inhibitor of the immunoproteasome subunit LMP7 blocks cytokine production and attenuates progression of experimental arthritis. *Nat Med* **15,** 781–787 (2009).

137. King, R. W., Deshaies, R. J., Peters, J. M. & Kirschner, M. W. How proteolysis drives the cell cycle. *Science* **274,** 1652–1659 (1996).

138. Adams, J. The proteasome: a suitable antineoplastic target. *Nat Rev Canc* **4,** 349–360 (2004).

139. Paramore, A. & Frantz, S. Bortezomib. *Nat Rev Drug Disc* **2,** 611–612 (2003).

140. Kisselev, A. F., van der Linden, W. A. & Overkleeft, H. S. Proteasome inhibitors: an expanding army attacking a unique target. *Chem Biol* **19,** 99–115 (2012).

141. Shoemaker, R. H. The NCI60 human tumour cell line anticancer drug screen. *Nat Rev Canc* **6,** 813–823 (2006).

142. Laatsch, H. Antibase Version 4.0—The Natural Compound Identifier. (2012).

143. Pommier, A., Pons, J. M. & Kocienski, P. J. An Asymmetric Synthesis of (-)-Tetrahydrolipstatin. *Synthesis* **12,** 1294–1300 (1994).

144. Hall, H. K., Jr. The nucleophile-initiated polymerization of α, α-disubstituted β-lactones. *Macromolecules* **2,** 488–497 (1969).

145. Riccio, R., Bifulco, G., Cimino, P., Bassarello, C. & Gomez-Paloma, L. Stereochemical analysis of natural products. Approaches relying on the combination of NMR spectroscopy and computational methods. *Pure Appl Chem* **75,** 295–308 (2003).

146. Hill, E. A. & Roberts, J. D. A Reinterpretation of the Nuclear Magnetic Resonance Spectrum of Cyclobutene. *J Am Chem Soc* **89,** 2047–2049 (1967).

147. Tello, R., Hallada, L. P., Niroula, D. & Rogeli, S. Total synthesis and absolute stereochemistry of the proteasome inhibitors cystargolides A and B. *Org Biomol Chem* **13,** 10127–10130 (2015).

148. Kawamura, S. *et al.* Potent proteasome inhibitors derived from the unnatural cis-cyclopropane isomer of belactosin A: synthesis, biological activity, and mode of action. *J Med Chem* **56,** 3689–3700 (2013).

149. Korotkov, V. S. *et al.* Synthesis and biological activity of optimized belactosin C congeners. *Org Biomol Chem* **9,** 7791–7798 (2011).

150. Nakamura, H., Watanabe, M., Ban, H. S., Nabeyama, W. & Asai, A. Synthesis and biological evaluation of boron peptide analogues of Belactosin C as proteasome inhibitors. *Bioorg Med Chem Lett* **19,** 3220–3224 (2009).

151. Finking, R. & Marahiel, M. A. Biosynthesis of nonribosomal peptides. *Ann Rev Microbiol* **58,** 453–488 (2004).

152. Strieker, M., Tanović, A. & Marahiel, M. A. Nonribosomal peptide synthetases: structures and dynamics. *Curr Op Struc Biol* **20,** 234–240 (2010).

153. Fischbach, M. A. & Walsh, C. T. Assembly-line enzymology for polyketide and nonribosomal Peptide antibiotics: logic, machinery, and mechanisms. *Chem Rev* **106,** 3468–3496 (2006).

154. Stachelhaus, T., Mootz, H. D. & Marahiel, M. A. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem Biol* **6,** 493–505 (1999).

155. Lambalot, R. H., Gehring, A. M., Flugel, R. S. & Zuber, P. A new enzyme superfamily—the phosphopantetheinyl transferases. *Chem Biol* **3,** 923–936 (1996).

156. Walsh, C. T. *et al.* Tailoring enzymes that modify nonribosomal peptides during and after chain elongation on NRPS assembly lines. *Curr Op Chem Biol* **5,** 525–534 (2001).

157. Chin, C.-S. *et al.* Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Meth* **10,** 563–569 (2013).

158. Aziz, R. K. *et al.* The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* **9,** 75–90 (2008).

159.Brettin, T. *et al.* RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci Rep* **5,** 8365–8371 (2015).

160. Overbeek, R. *et al.* The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucl Acids Res* **42,** 206–214 (2014).

161. Weber, T. *et al.* antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucl Acids Res* **43,** 237–244 (2015).

162. Cobb, R. E., Wang, Y. & Zhao, H. High-efficiency multiplex genome editing of *Streptomyces* species using an engineered CRISPR/Cas system. *ACS Synth Biol* **4,** 723–728 (2014).

163. Kieser, T. *Practical Streptomyces Genetics*. 162–170 (John Innes Foundation, 2000).

164. Engler, C., Kandzia, R. & Marillonnet, S. A one pot, one step, precision cloning method with high throughput capability. *PLoS One* **3,** e3647 (2008).

165. Stackebrandt, E. & Goodfellow, M. *Nucleic Acid Techniques in Bacterial Systematics*. **120,** 115–175 (John Wiley & Son Ltd, 1991).

166. Lim, S.-K. *et al.* iso-Migrastatin, migrastatin, and dorrigocin production in *Streptomyces platensis* NRRL 18993 is governed by a single biosynthetic machinery featuring an acyltransferase-less type I polyketide synthase. *J Biol Chem* **284,** 29746–29756 (2009).

167. Barona-Gómez, F., Wong, U., Giannakopulos, A. E., Derrick, P. J. & Challis, G. L. Identification of a cluster of genes that directs desferrioxamine biosynthesis in *Streptomyces coelicolor* M145. *J Am Chem Soc* **126,** 16282–16283 (2004).

168. Hoshino, T. The cyclization mechanism of squalene in hopene biosynthesis: the terminal methyl groups are critical to the correct folding of this substrate both for the formation of the five-membered E-ring and for the initiation of the polycyclization reaction. *Chem Comm* **8,** 731–732 (1999).

169. Röttig, M. *et al.* NRPSpredictor2--a web server for predicting NRPS adenylation domain specificity. *Nucl Acids Res* **39,** 362–367 (2011).

170. Zhang, W., Heemstra, J. R., Jr, Walsh, C. T. & Imker, H. J. Activation of the pacidamycin PacL adenylation domain by MbtH-like proteins. *Biochemistry* **49,** 9946–9947 (2010).

171. Keating, T. A. *et al.* Chain termination steps in nonribosomal peptide synthetase assembly lines: directed acyl-S-enzyme breakdown in antibiotic and siderophore biosynthesis. *Chembiochem* **2,** 99–107 (2002).

172. Kawaguchi, A., Iwamoto, A. & Sato, N. Comprehensive natural products chemistry. *Comprehensive Natural Products* **1,** 23–59 (Elsevier Science Ltd., 1999).

173. Kopp, F., Linne, U. & Oberthür, M. Harnessing the chemical activation inherent to carrier protein-bound thioesters for the characterization of lipopeptide fatty acid tailoring enzymes. *J Am Chem Soc* **130,** 2656 (2008).

174. Hubbard, B. K., Thomas, M. G. & Walsh, C. T. Biosynthesis of Lp-hydroxyphenylglycine, a non-proteinogenic amino acid constituent of peptide antibiotics. *Chem Biol* **12,** 931–942 (2000).

175. Alkhalaf, L. M. & Ryan, K. S. Biosynthetic manipulation of tryptophan in bacteria: pathways and mechanisms. *Chem Biol* **22,** 317–328 (2015).

176. Mockus, S. M. & Vrana, K. E. Advances in the molecular characterization of tryptophan hydroxylase. *J Mol Neurosci* **10,** 163–179 (1998).

177. Crosa, J. H. & Walsh, C. T. Genetics and assembly line enzymology of siderophore biosynthesis in bacteria. *Microbiol Mol Bio Rev* **66,** 223–249 (2002).

178. Wang, H., Fewer, D. P., Holm, L., Rouhiainen, L. & Sivonen, K. Atlas of nonribosomal peptide and polyketide biosynthetic pathways reveals common occurrence of nonmodular enzymes. *PNAS* **111,** 9259–9264 (2014).

179. Christiansen, G., Philmus, B., Hemscheidt, T. & Kurmayer, R. Genetic variation of adenylation domains of the anabaenopeptin synthesis operon and evolution of substrate promiscuity. *J Bacteriol* **193,** 3822–3831 (2011).

180. Wyatt, M. A. *et al.* Biosynthesis of ebelactone A: isotopic tracer, advanced precursor and genetic studies reveal a thioesterase-independent cyclization to give a polyketide β-lactone. *J Antibiot* **66,** 421–430 (2013).

181. Udwary, D. W. *et al.* Genome sequencing reveals complex secondary metabolome in the marine actinomycete *Salinispora tropica*. *PNAS* **104,** 10376–10381 (2007).

182. Bai, T. *et al.* Operon for biosynthesis of lipstatin, the beta-lactone inhibitor of human pancreatic lipase. *Appl Environ Microbiol* **80,** 7473–7483 (2014).

183. Zhao, C. *et al.* Oxazolomycin biosynthesis in *Streptomyces albus* JA3453 featuring an ‘acyltransferase-less’ type I polyketide synthase that incorporates two distinct extender units. *J Biol Chem* **285,** 20097–20108 (2010).

184. Beer, L. L. & Moore, B. S. Biosynthetic convergence of salinosporamides A and B in the marine actinomycete *Salinispora tropica*. *Org Lett* **9,** 845–848 (2007).

185. Harper, A. E., Miller, R. H. & Block, K. P. Branched-chain amino acid metabolism. *Ann Rev Nut* **4,** 409–454 (1984).

APPENDIX

8.1 NMR and MN^n Spectra of Known Compounds Isolated from RKKD and *Kitasatospora* Libraries

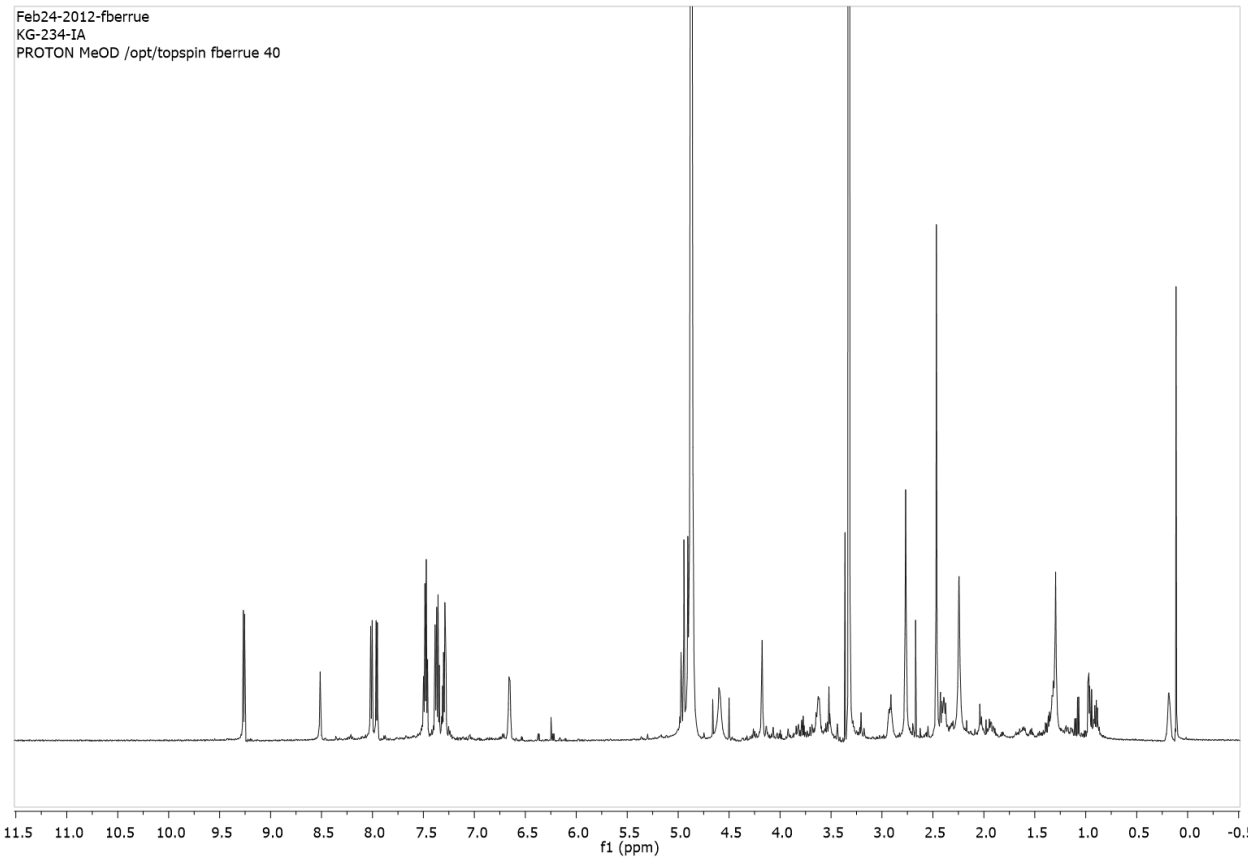


Figure 8.1. ^{1}H NMR spectrum of RKKD234_467, staurosporine.

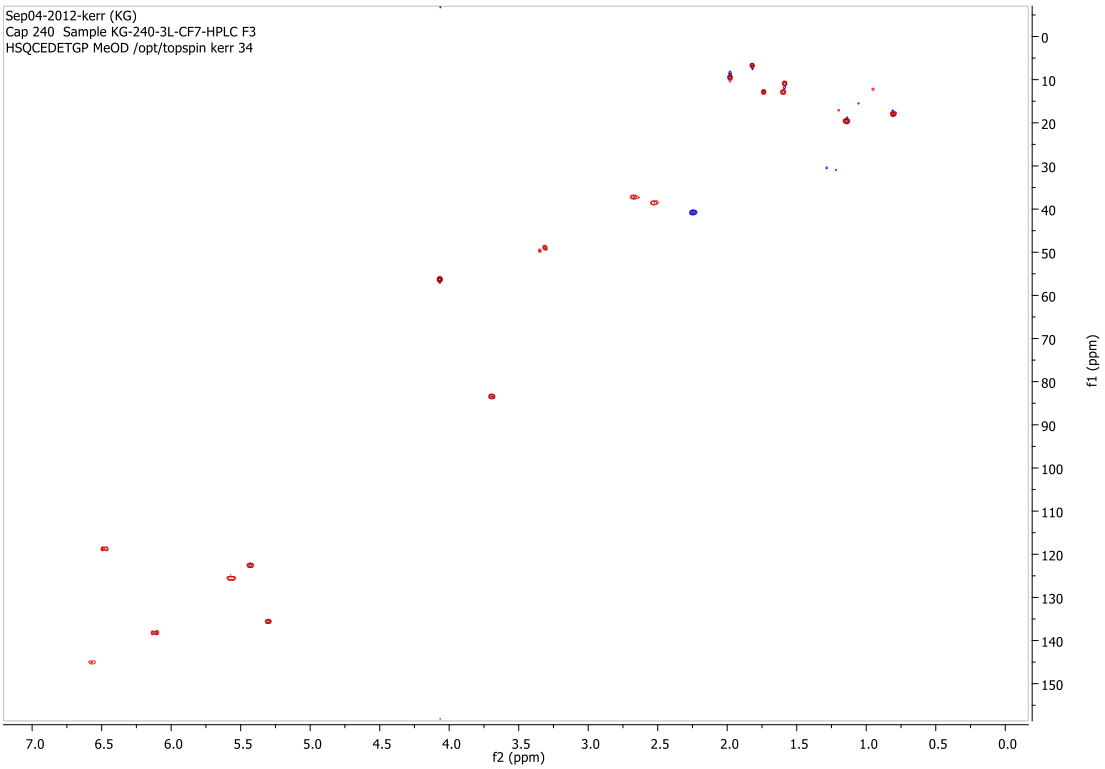


Figure 8.2. HSQC spectrum of RKKD240_401, PM05046.

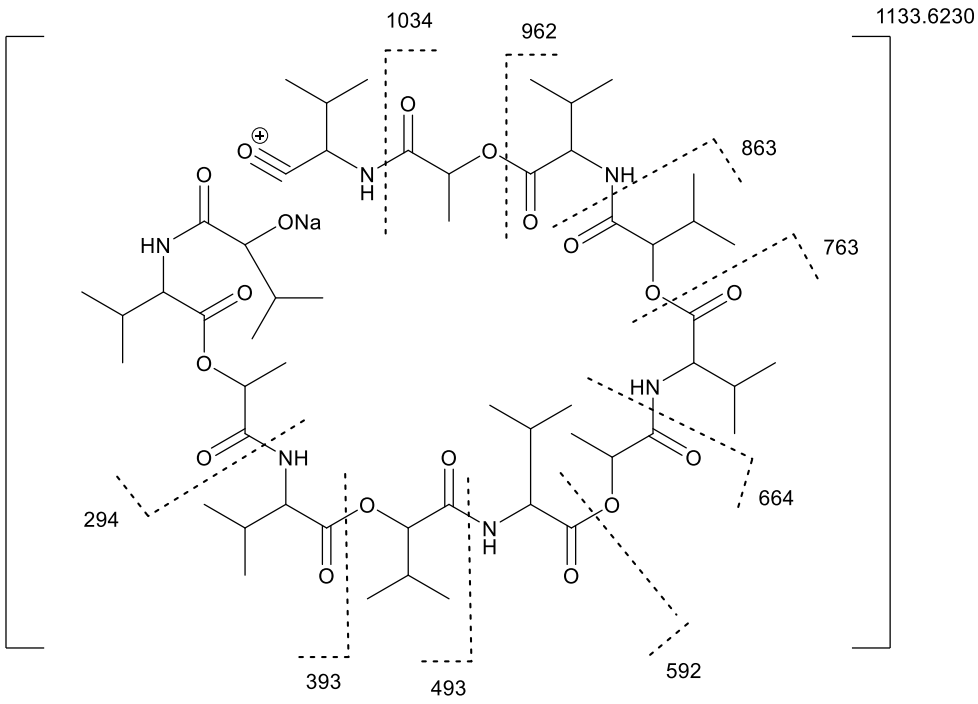


Figure 8.3. Summary of MS^n fragments used to determine that the structure of RKKD790_1128 was valinomycin.

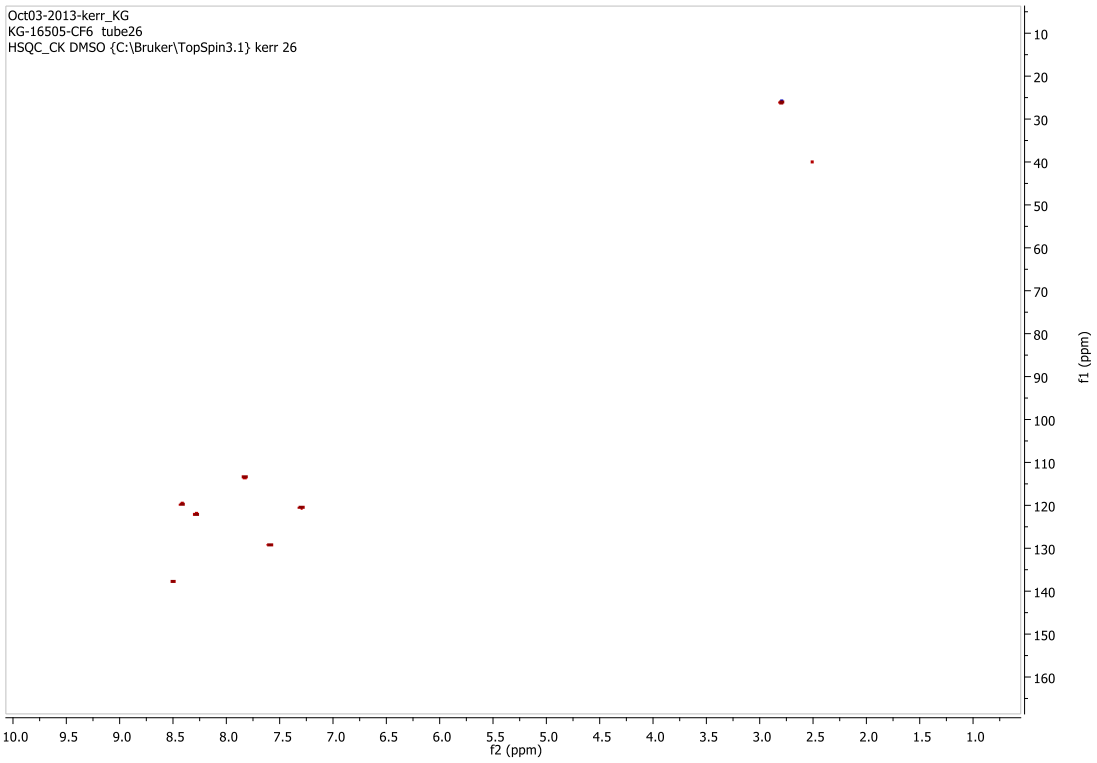


Figure 8.4. HSQC spectrum of 16505_211, 1-acetyl β-carboline.

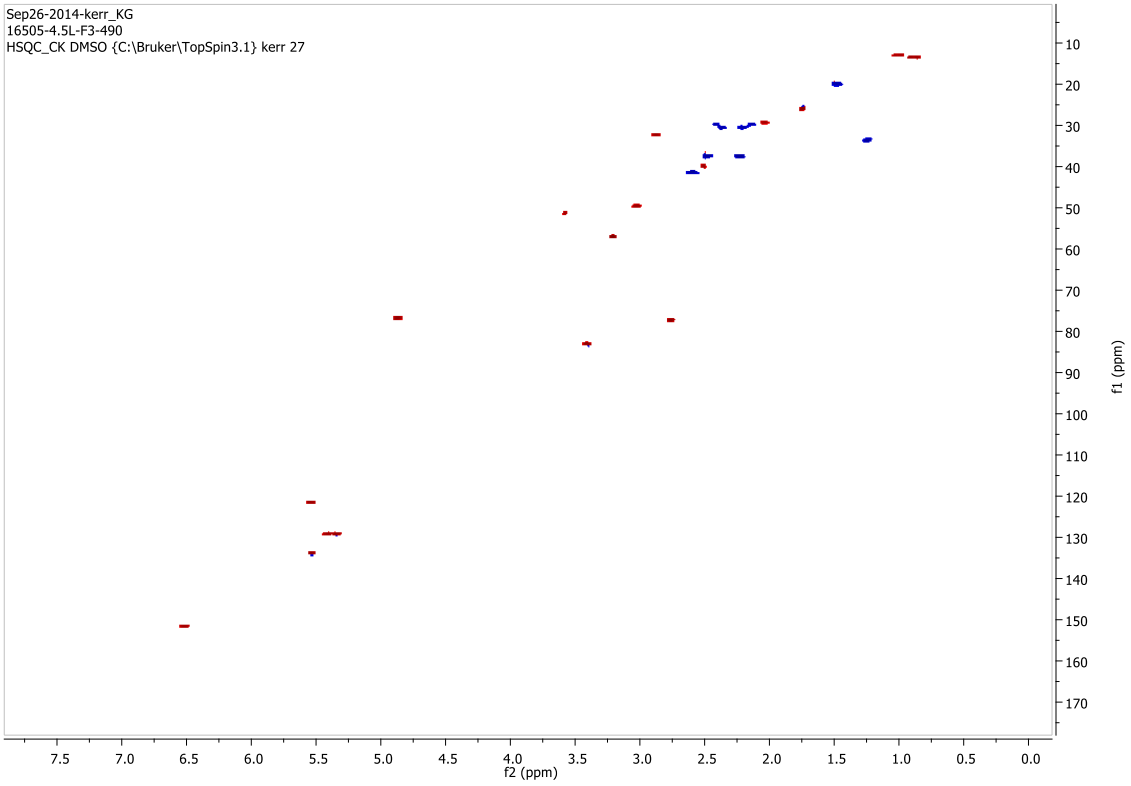


Figure 8.5. HSQC spectrum of 16505_490, migrastatin.

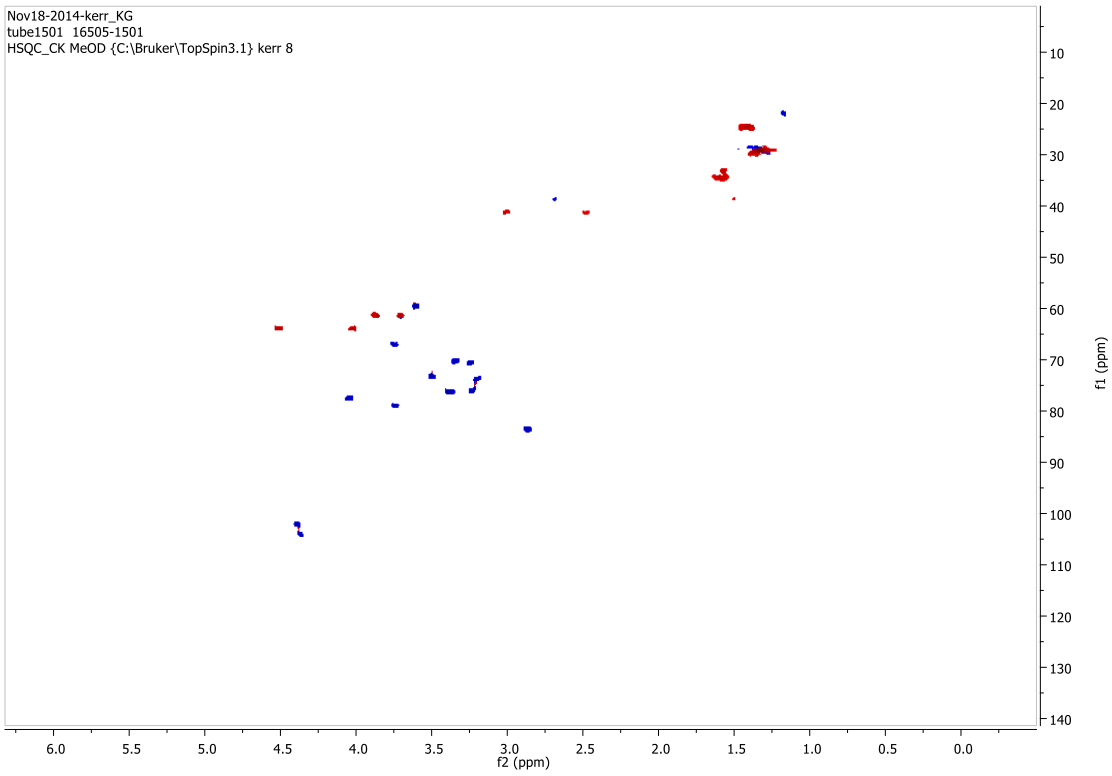


Figure 8.6. HSQC spectrum of 16505_1501, fattiviracin FV13.

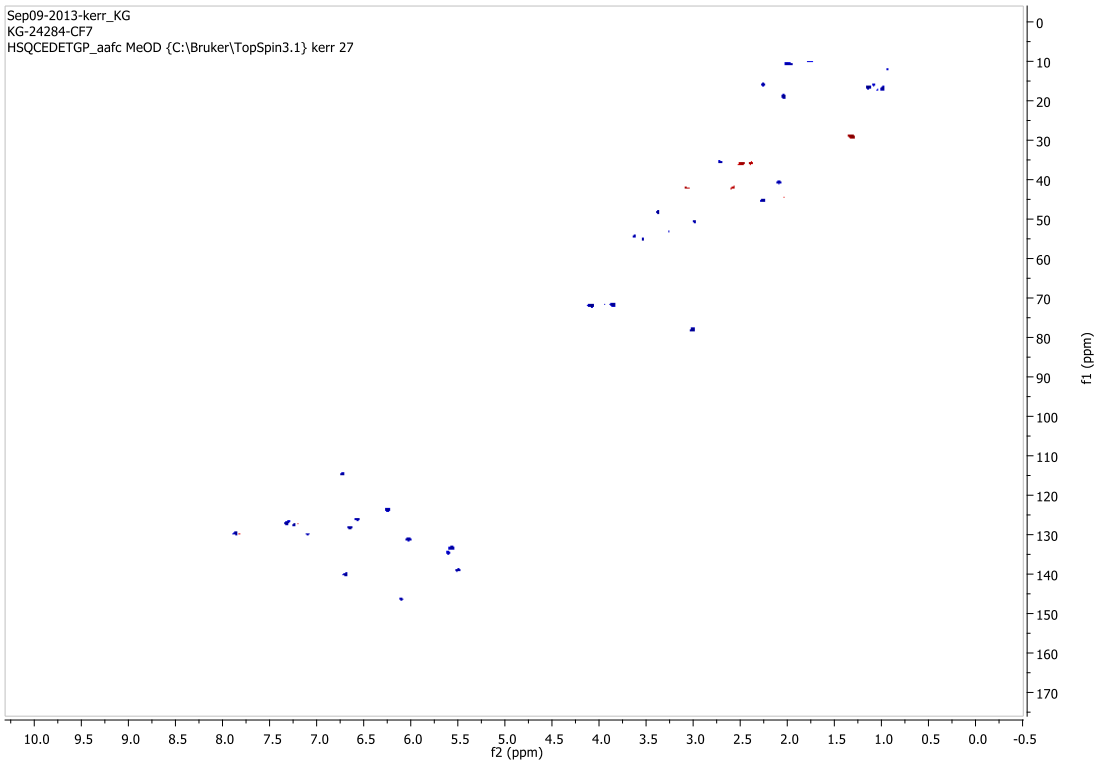


Figure 8.7. HSQC spectrum of 24284_720, naphthomycin A

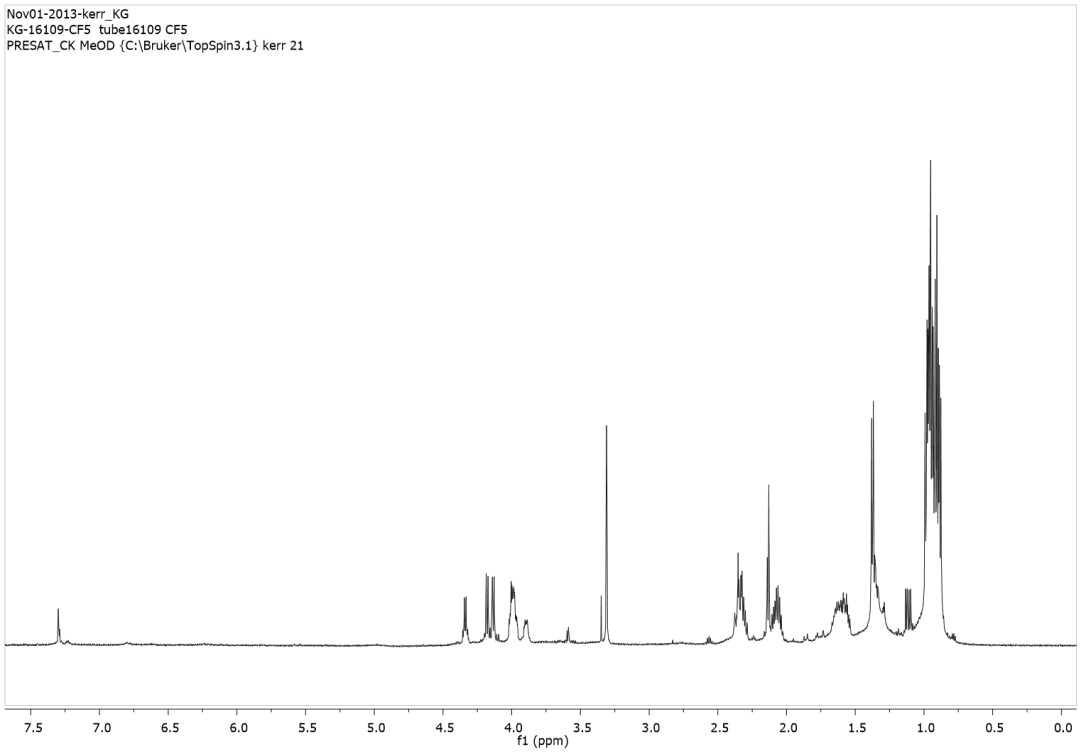
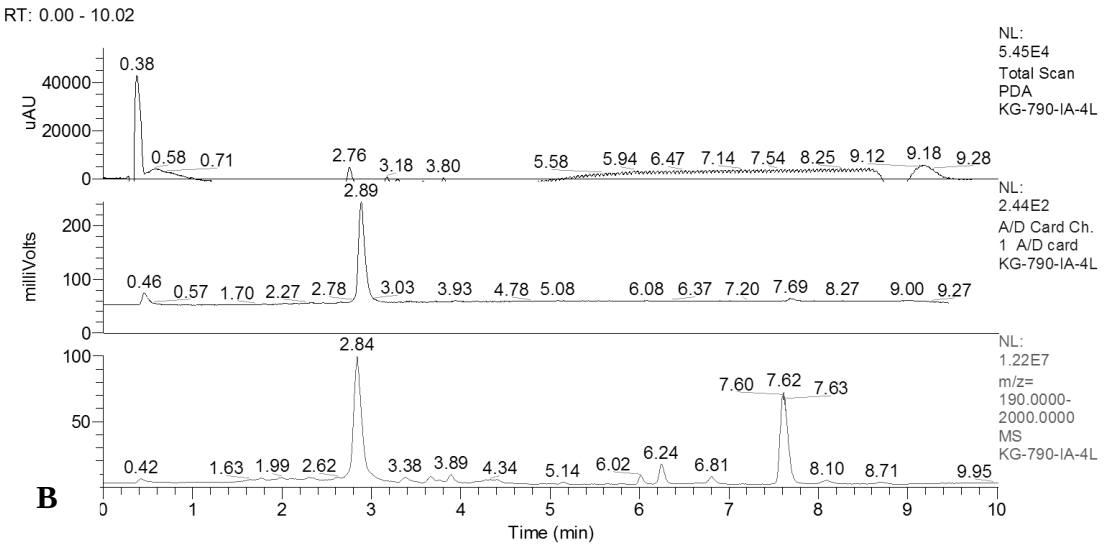


Figure 8.8. ^{1}H NMR spectrum of 16109_686, pepstatin A.

8.2 NMR and MS Spectra Used to Elucidate the Structure of RKKD790_1592

A



B

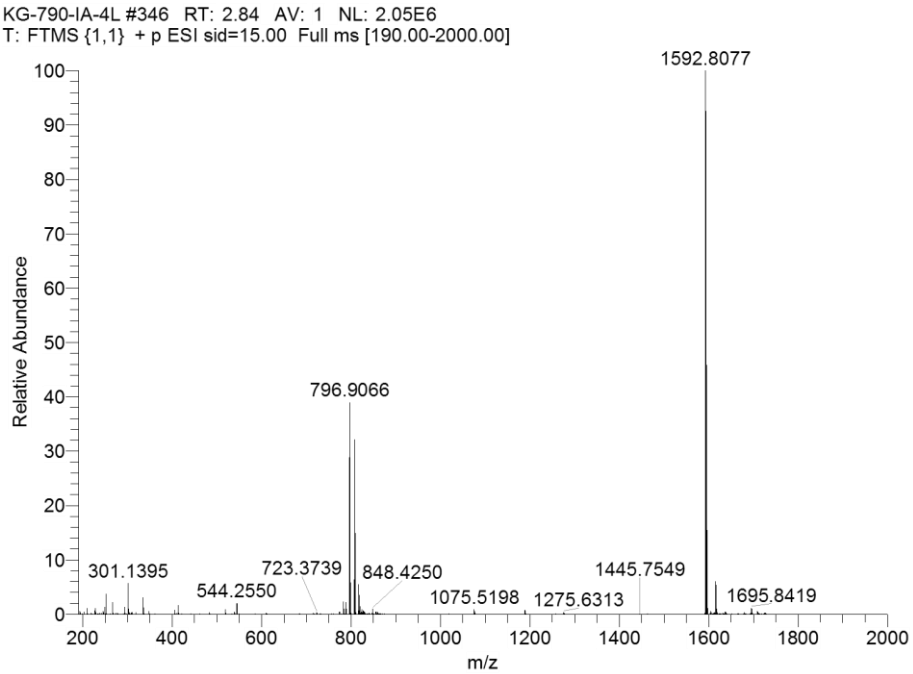


Figure 8.9. Chromatograms of the crude extract from a large scale fermentation of RKKD-790 grown in ISP2.
A. Three detectors were used to generate 3 chromatograms for the same sample: PDA, ELSD and +ESI-MS. **B**. The high resolution mass spectrum of RKKD790_1592.

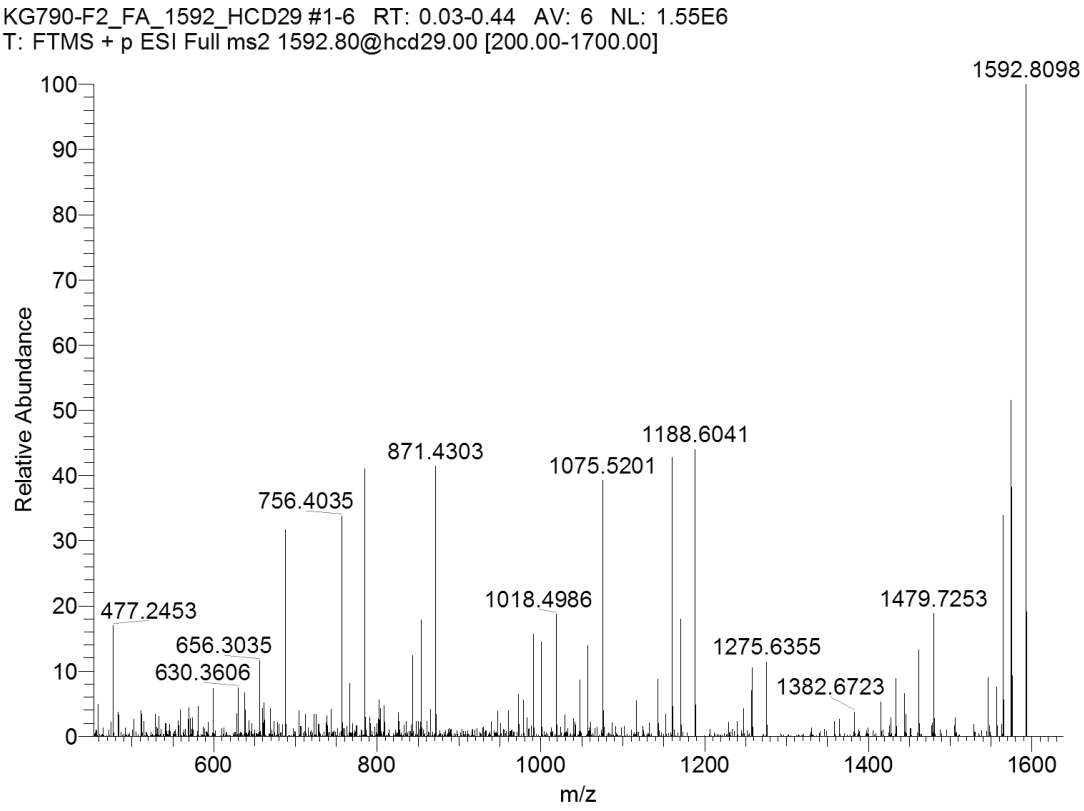


Figure 8.10. The MS^2 spectrum of RKKD790_1592 that was used to confirm the sequence of amino acids, and the lasso conformation.

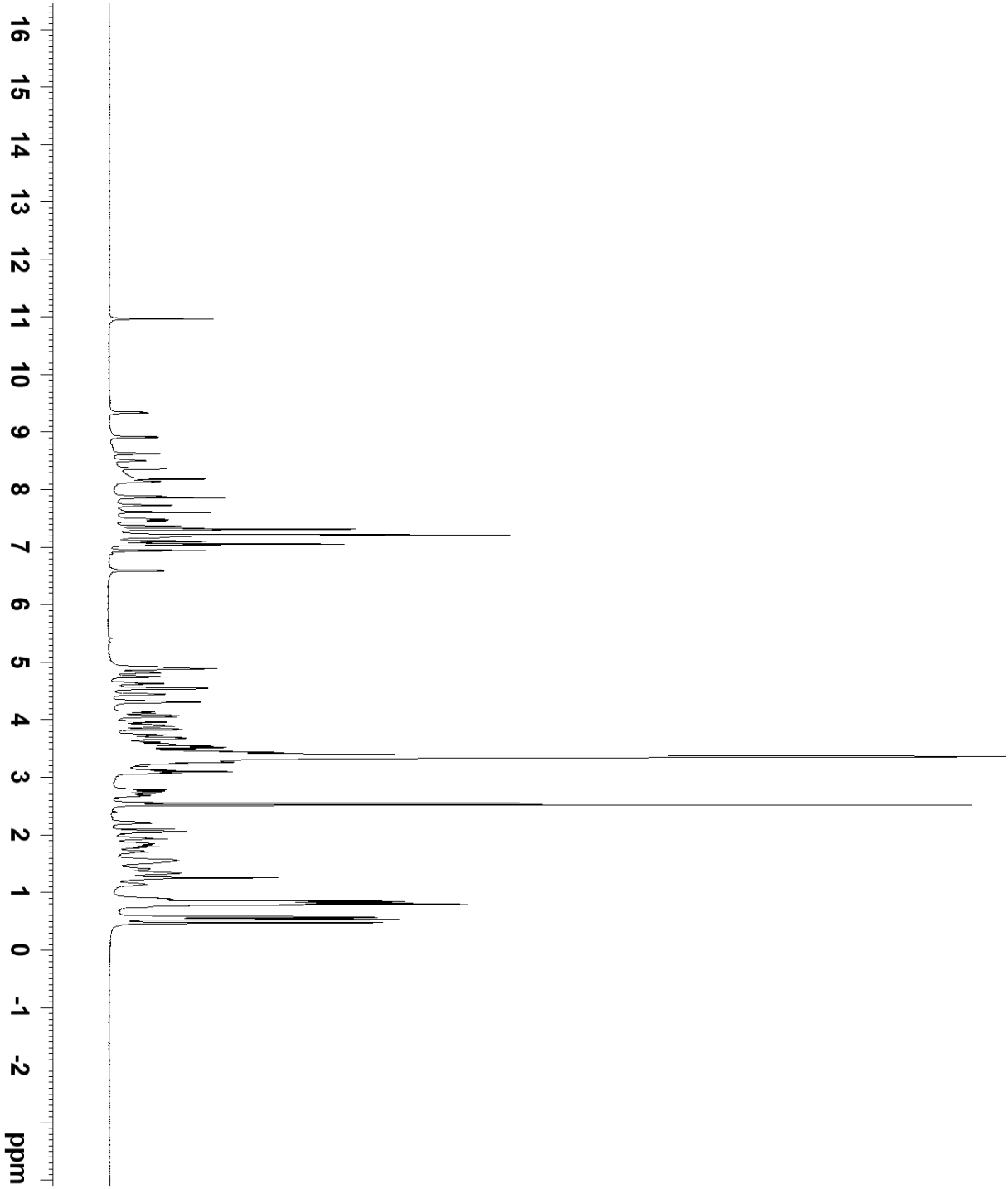


Figure 8.11. ^{1}H NMR (600MHz, DMSO-d_6) spectrum of RKKD790_1592.

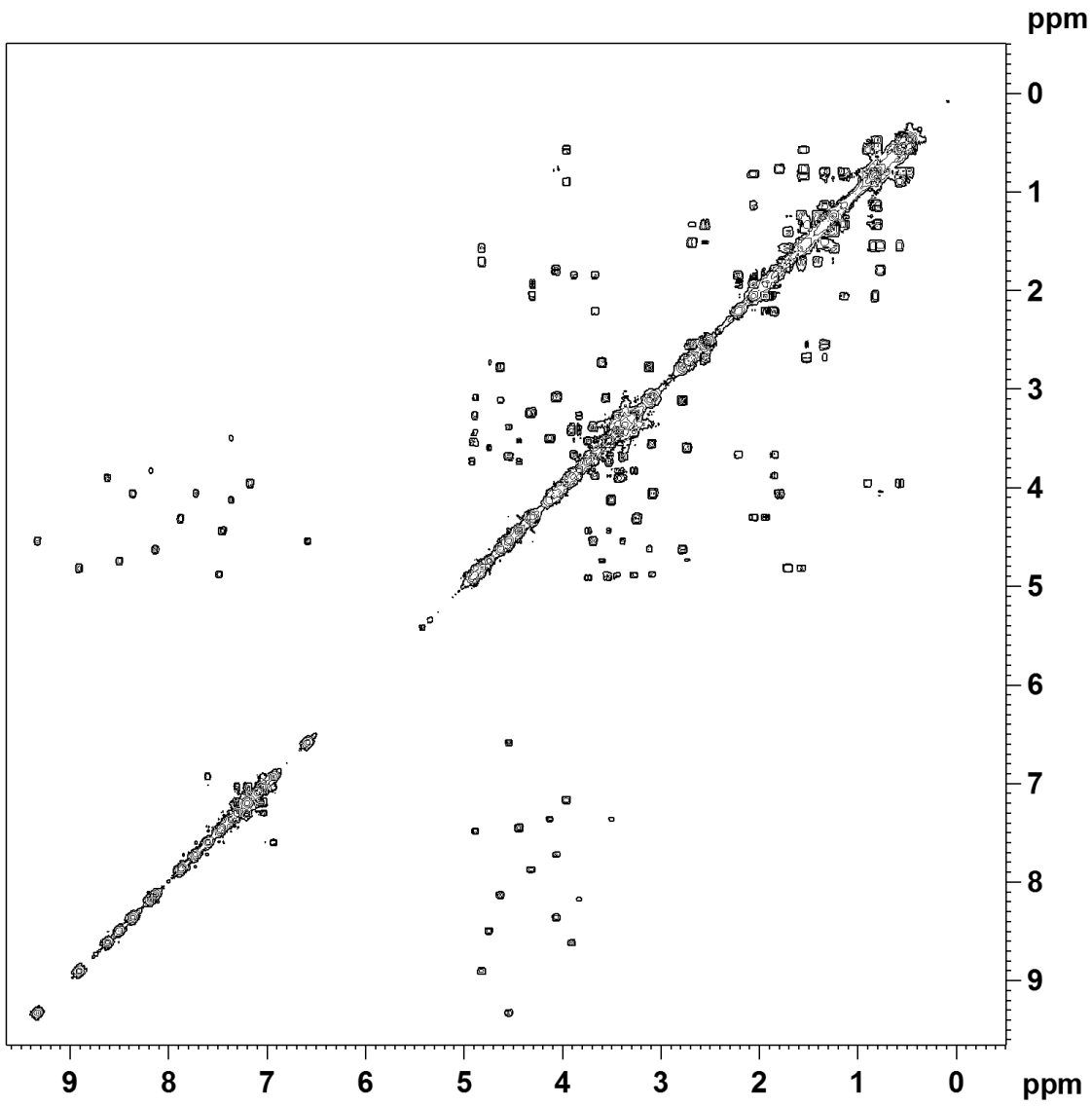


Figure 8.12. The COSY NMR spectrum of RKKD790_1592.

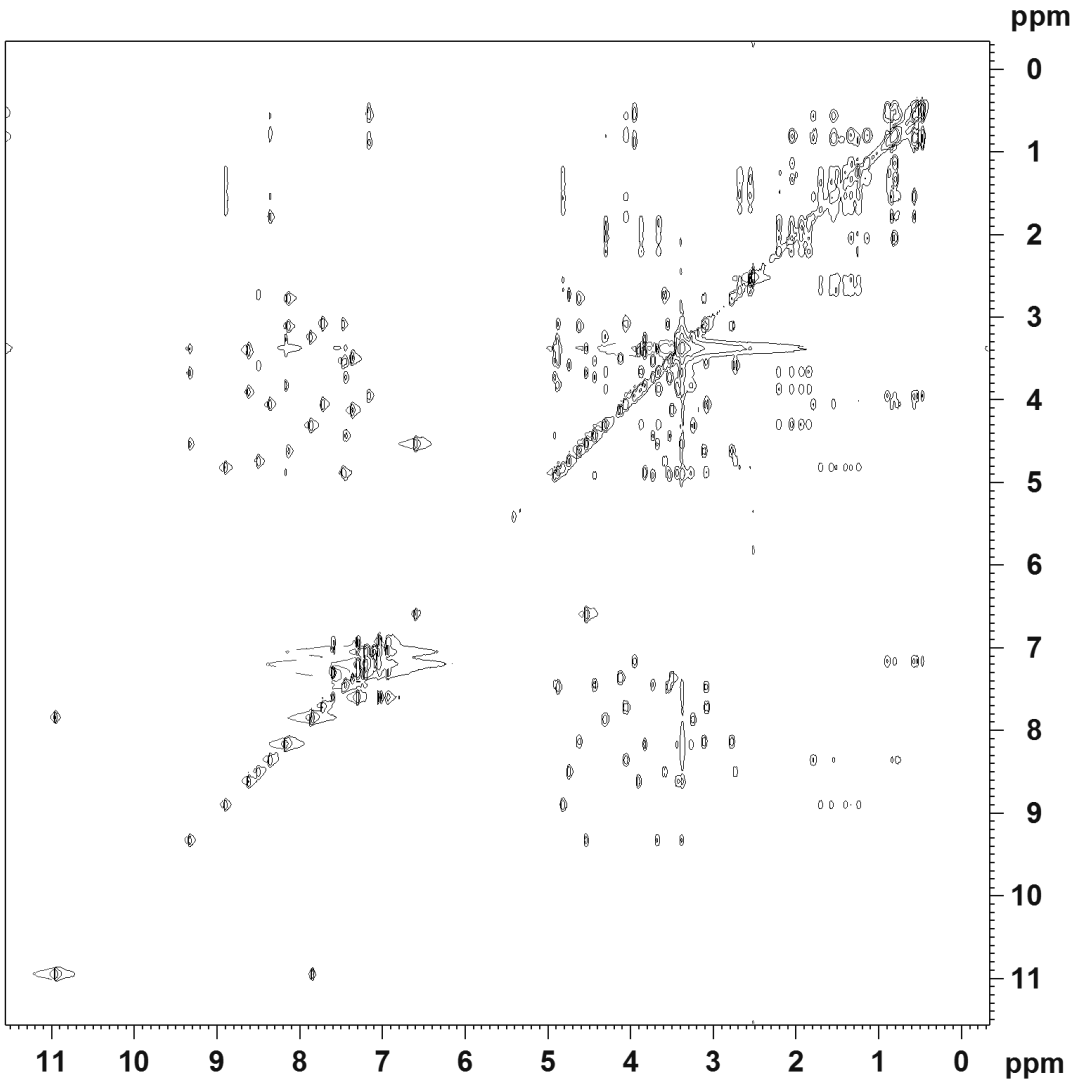


Figure 8.13. The TOCSY NMR spectrum of RKKD790_1592.

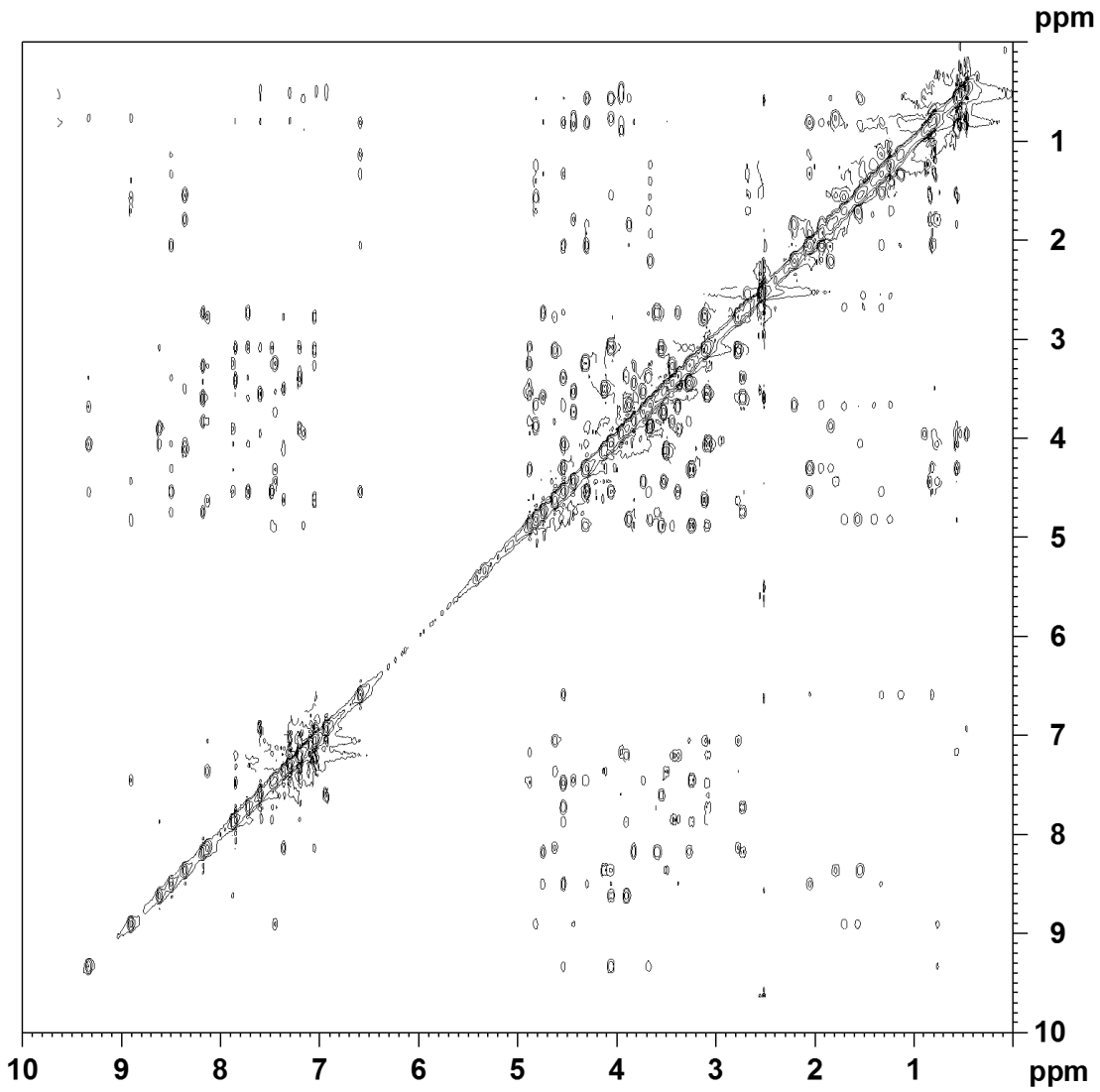


Figure 8.14. The ROESY NMR spectrum of RKKD790_1592.

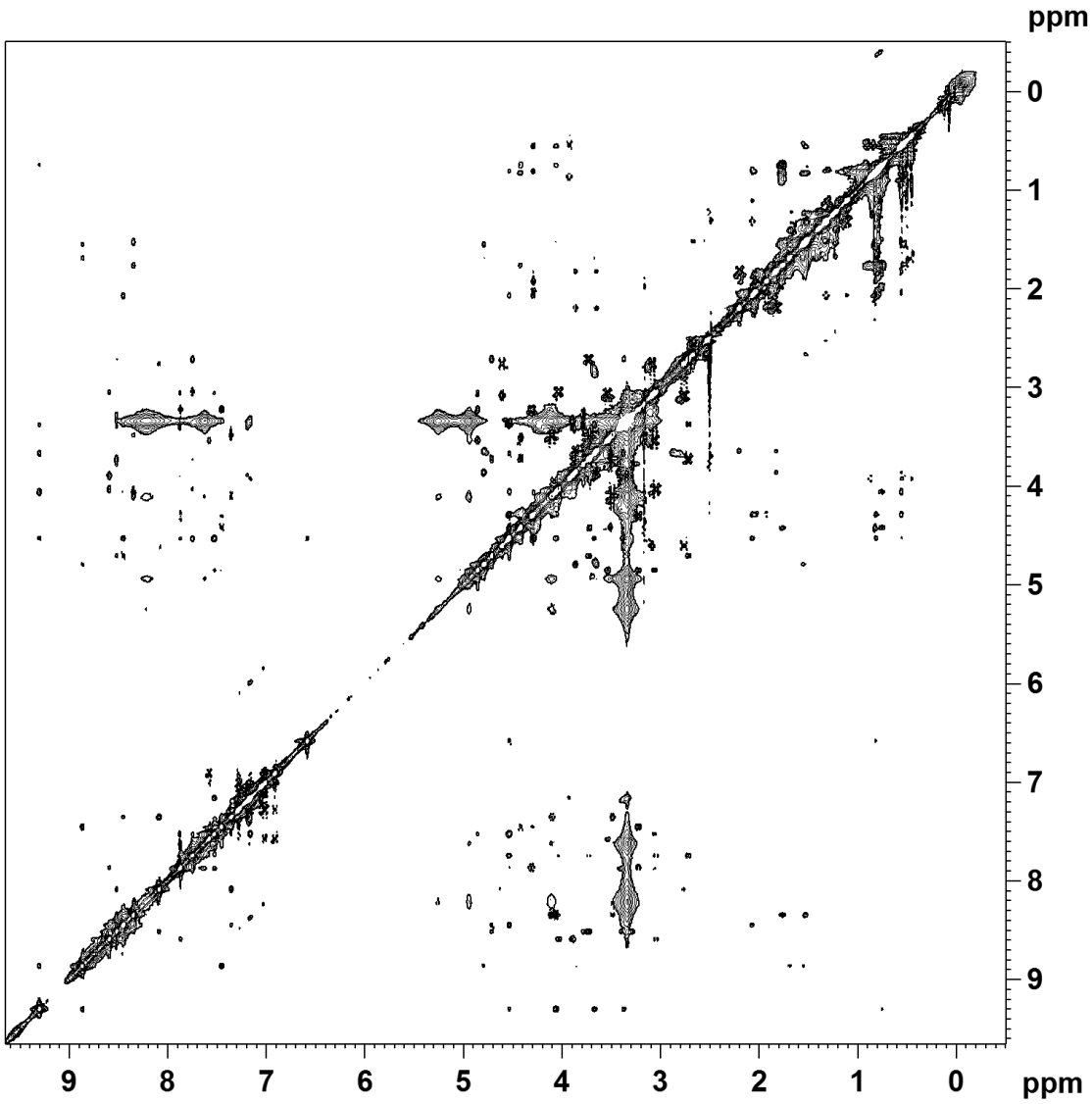


Figure 8.15. The NOESY NMR spectrum of RKKD790_1592.

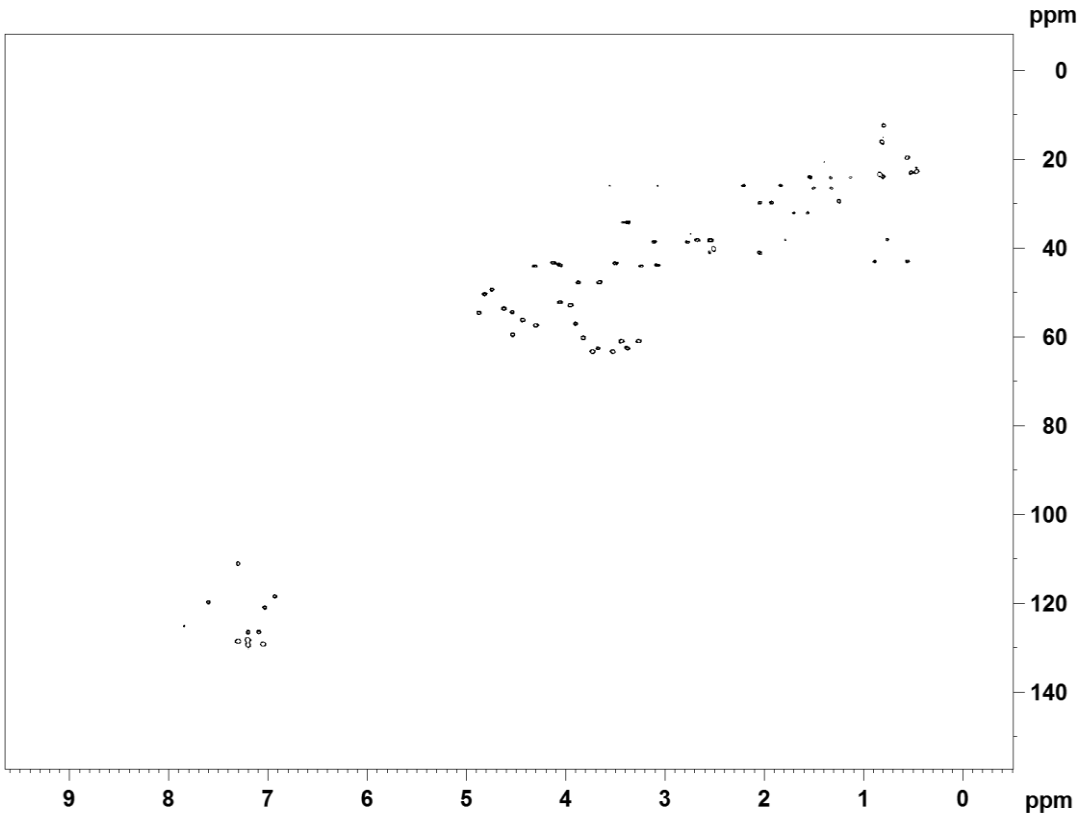


Figure 8.16. The HSQC NMR spectrum of RKKD790_1592.

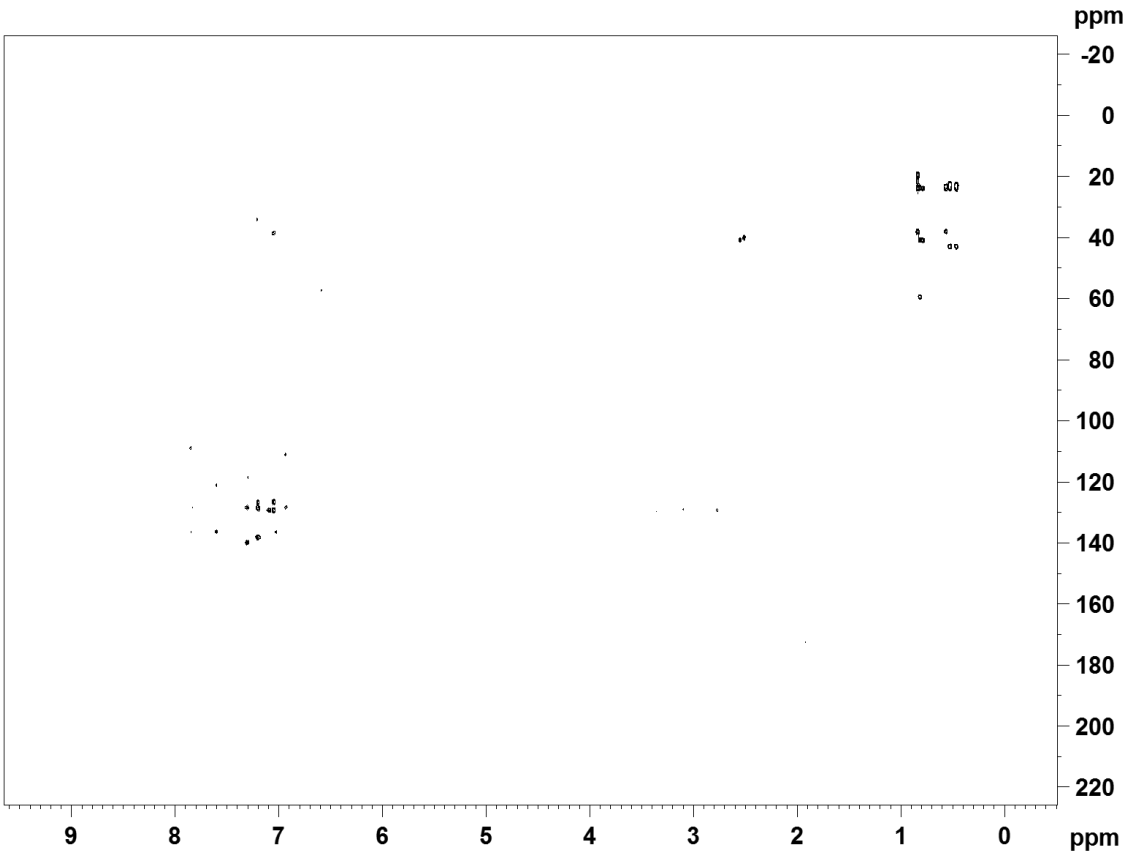


Figure 8.17. The HMBC NMR spectrum of RKKD790_1592.

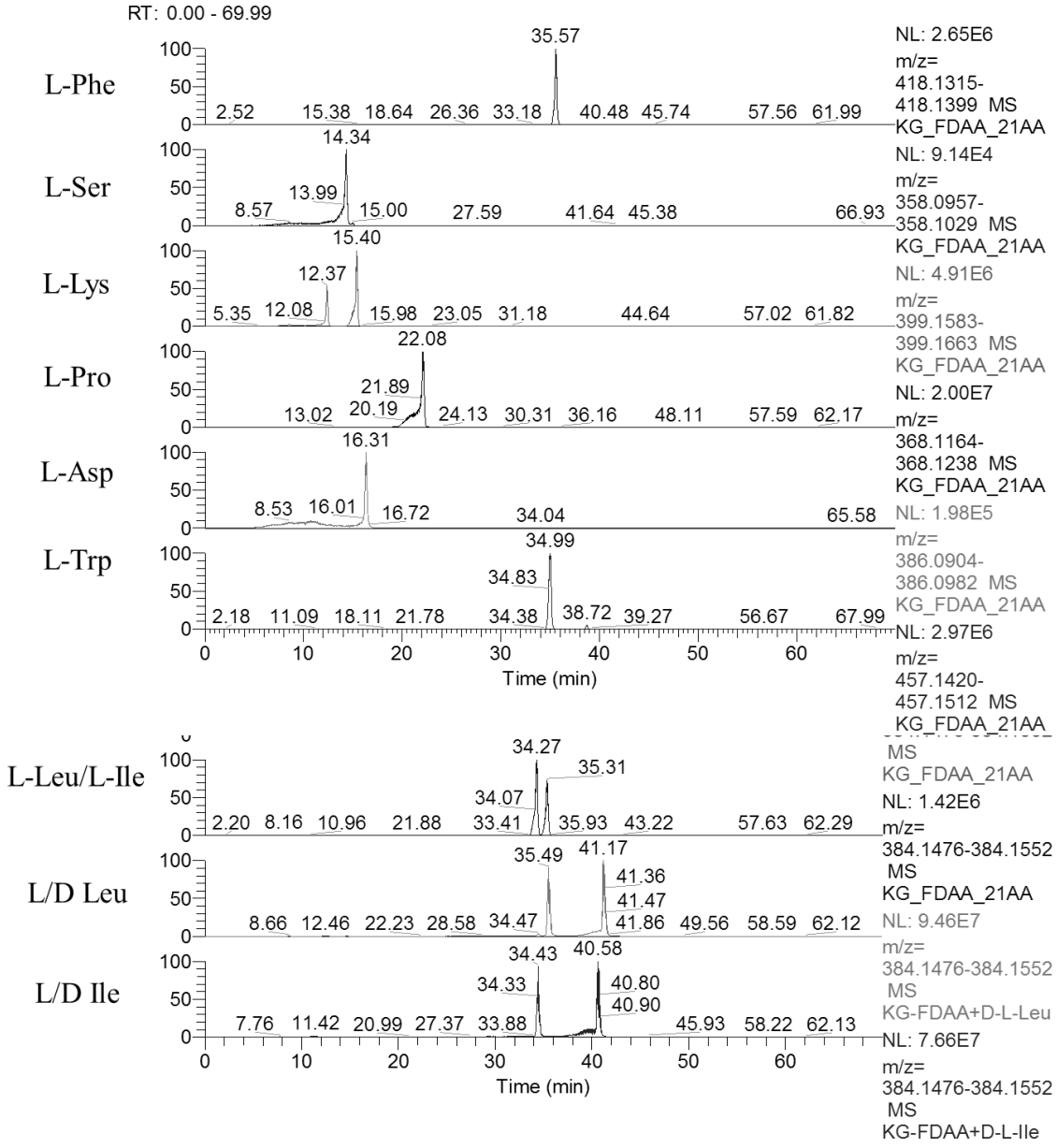


Figure 8.18. Marfey's Analysis to determine the absolute stereochemistry of amino acids. SIM +ESIMS chromatograms of derivatized amino acid standards.

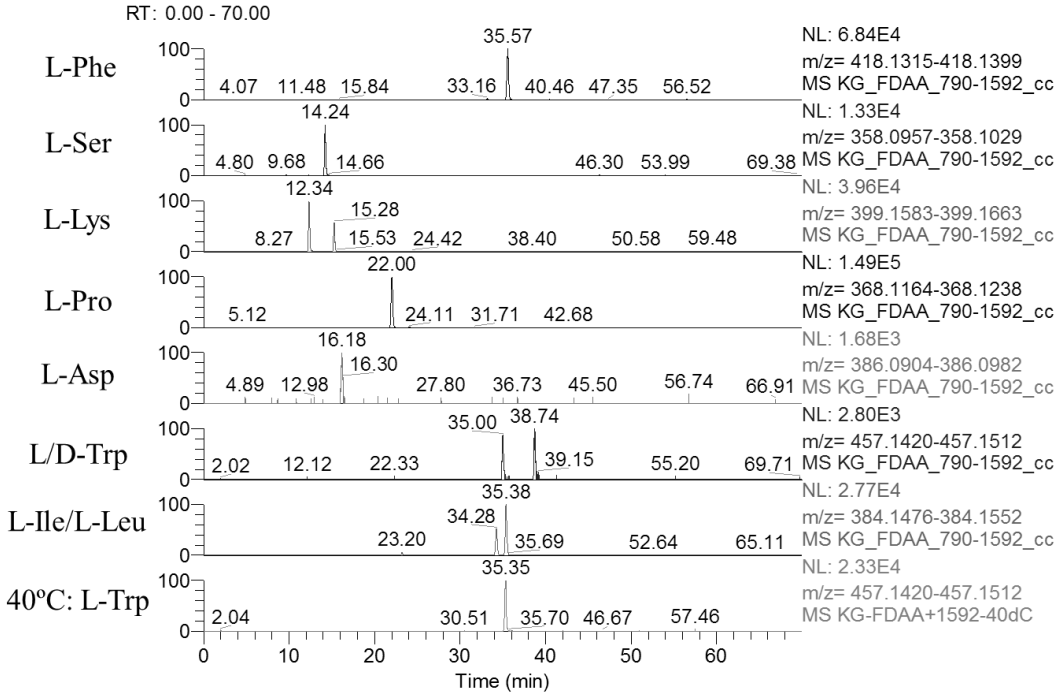


Figure 8.19. Marfey's Analysis of RKKD790_1592. SIM +ESIMS chromatograms of derivatized RKKD790_1592.

8.3 NMR and MS Spectra Used to Elucidate the Structure of Cystargamide

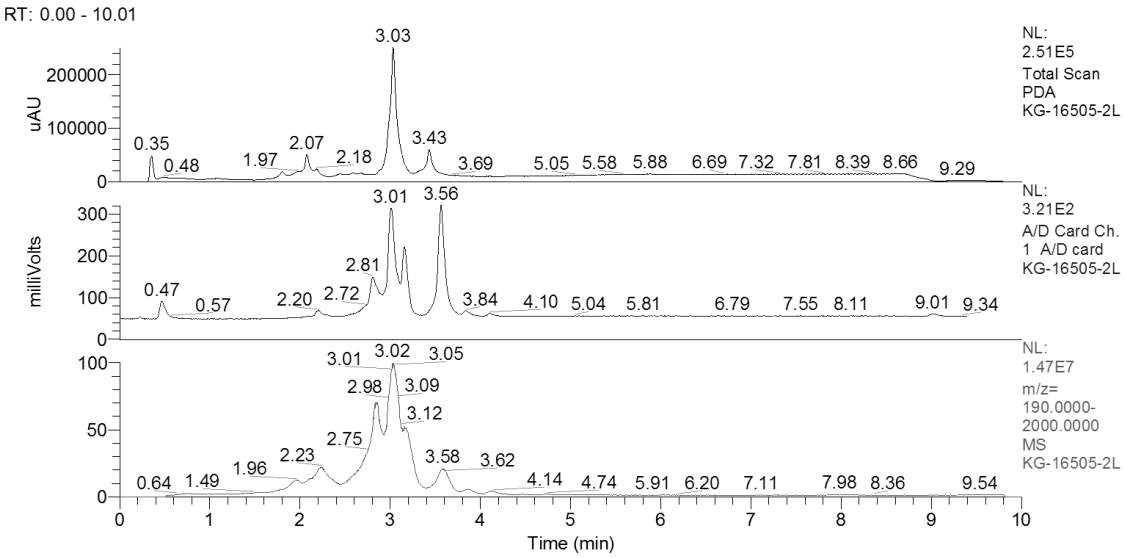


Figure 8.20. Chromatograms of the crude extract from a large scale fermentation of *Kitasatospora cystarginea* grown in 1045 medium.
Three detectors were used to generate 3 chromatograms for the same sample: PDA, ELSD and +ESI-MS.

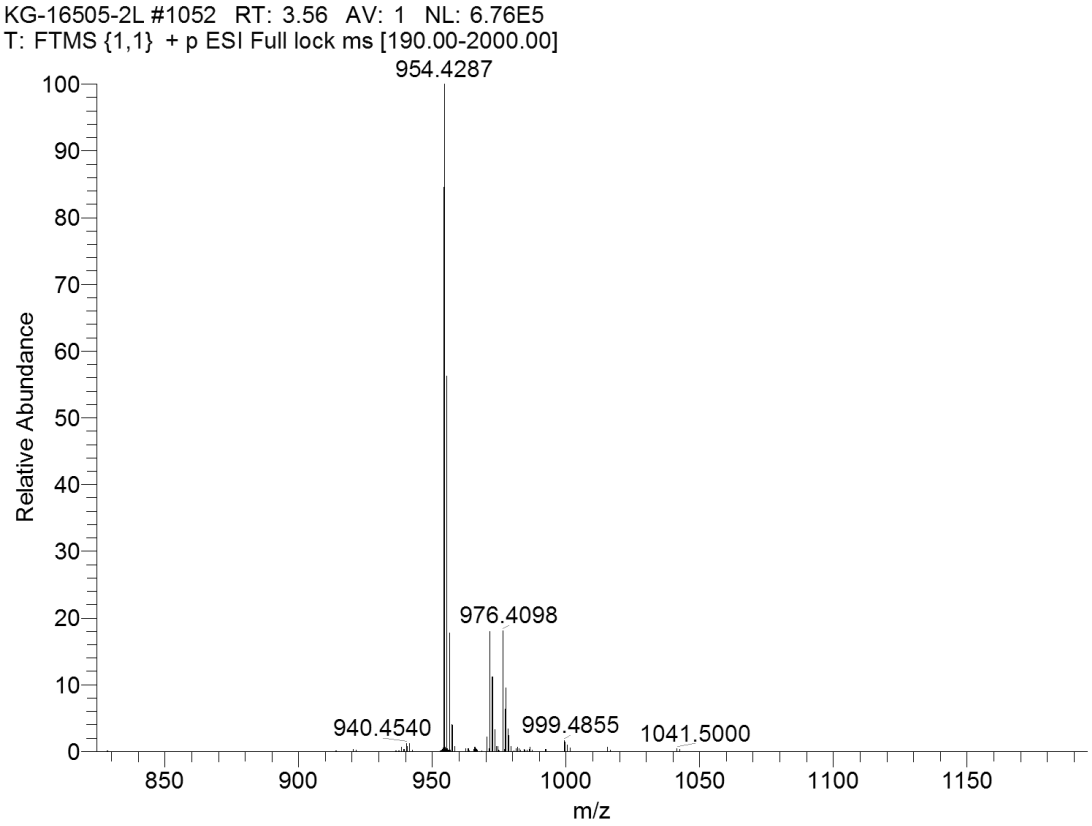


Figure 8.21. Figure 4.7. +ESI HRMS of cystargamide.

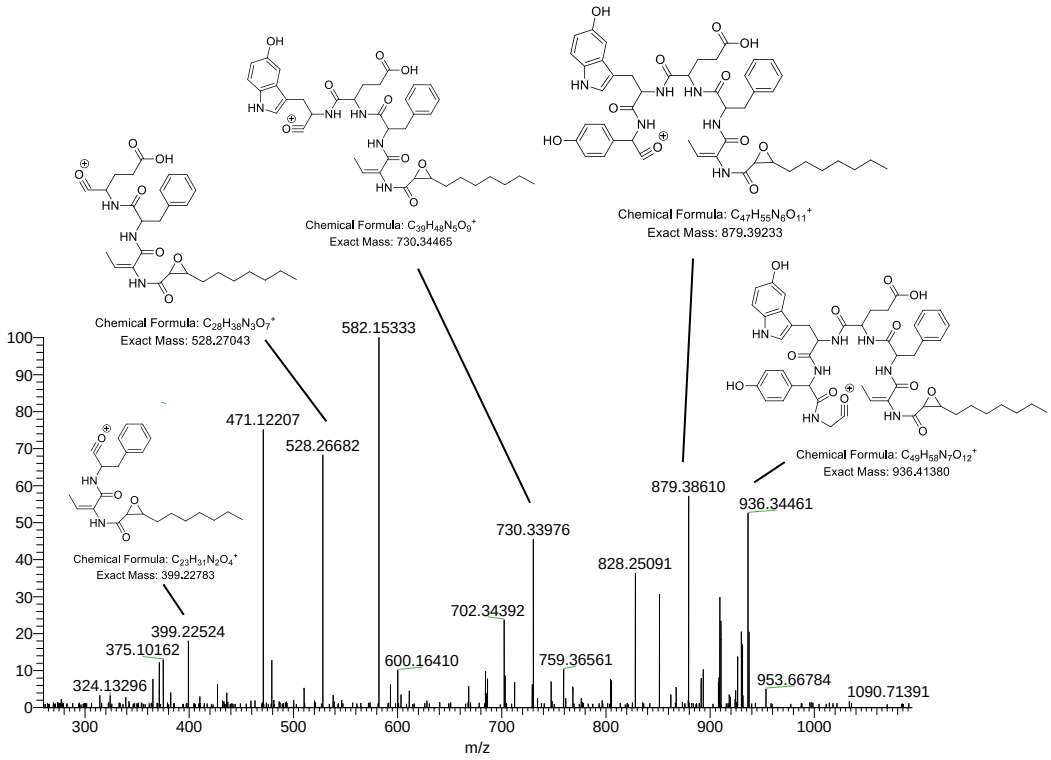


Figure 8.22. MS^2 spectrum of cystargamide.

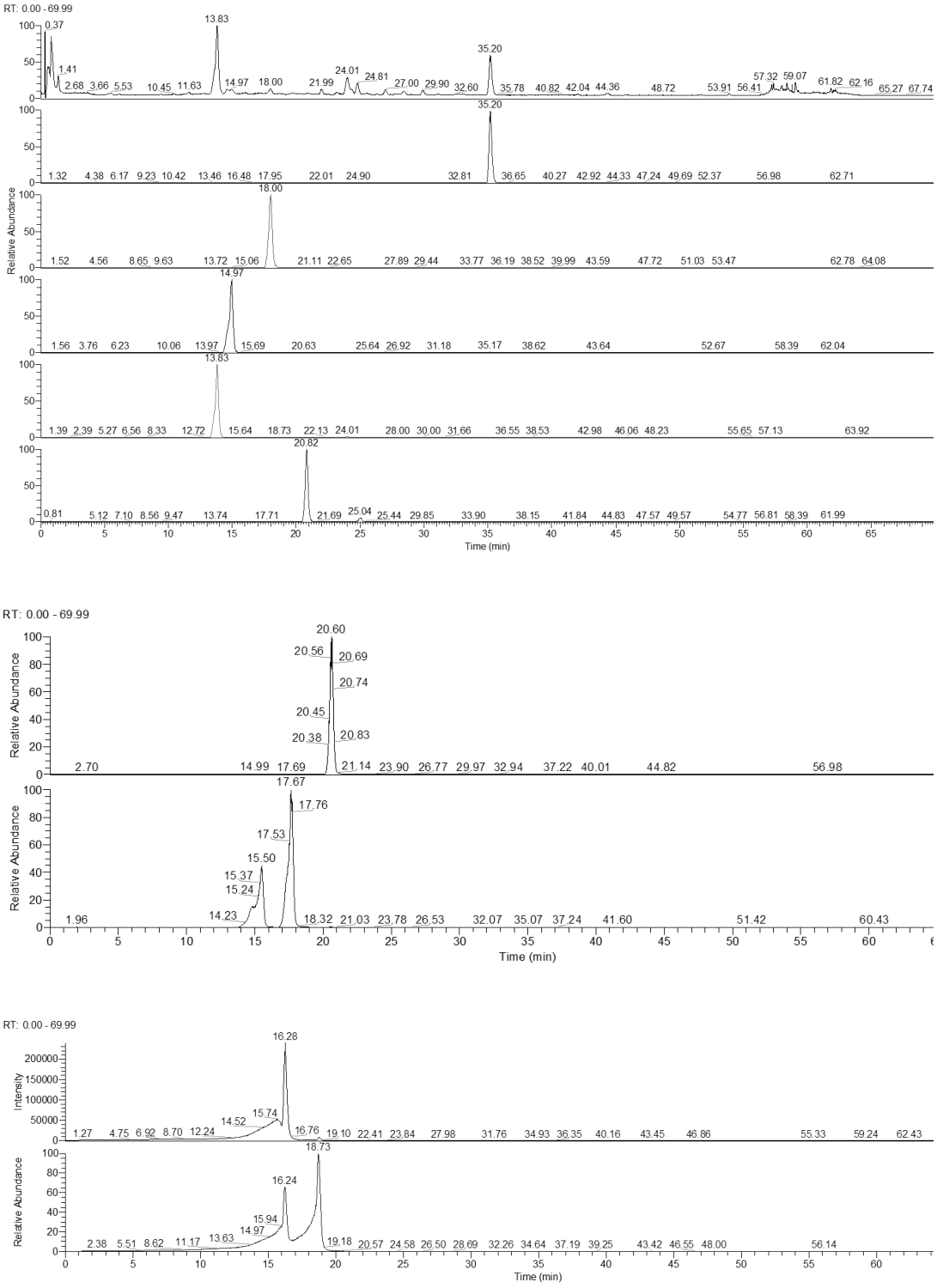


Figure 8.23. Chromatograms of L-FDAA derivatized amino acid standards.

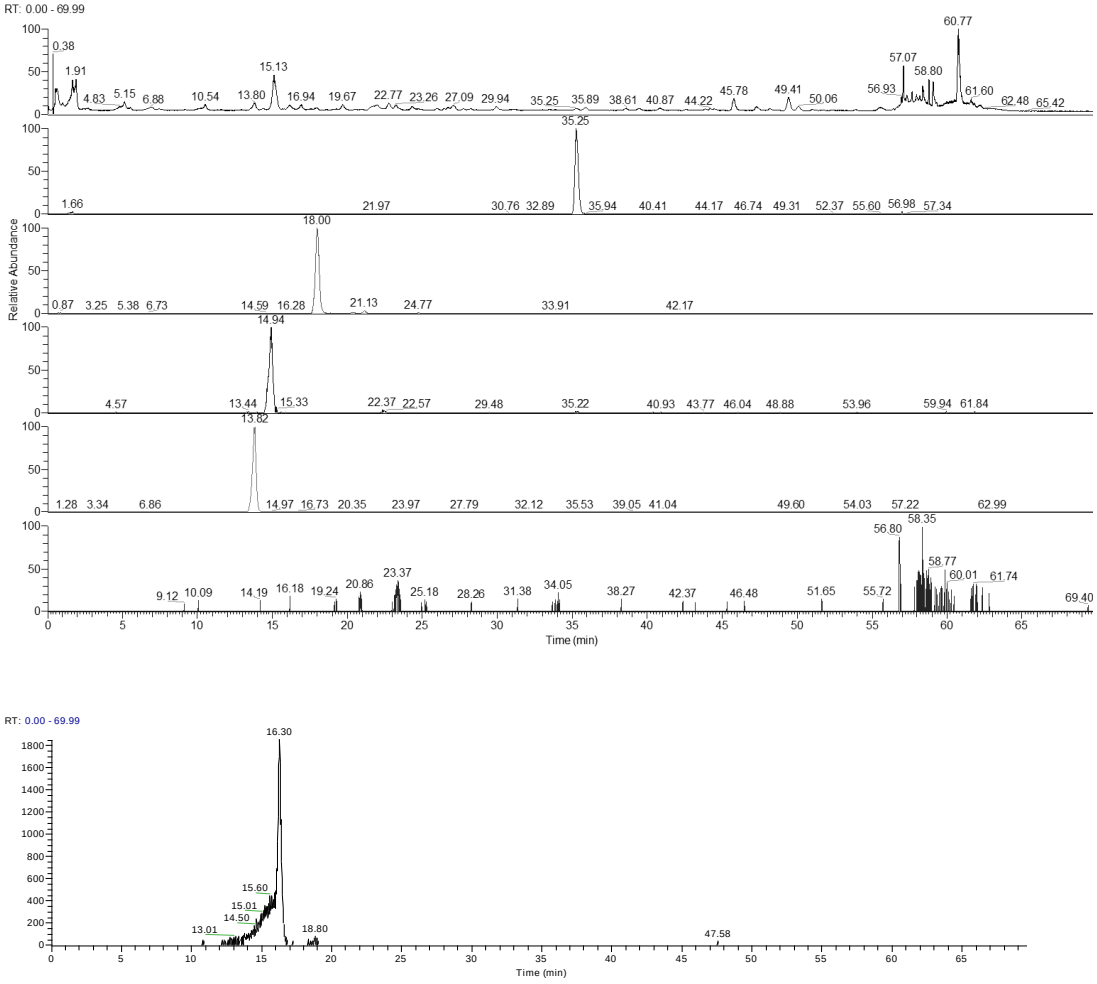


Figure 8.24. Chromatograms of L-FDAA derivatized hydrolysate of cystargamide.

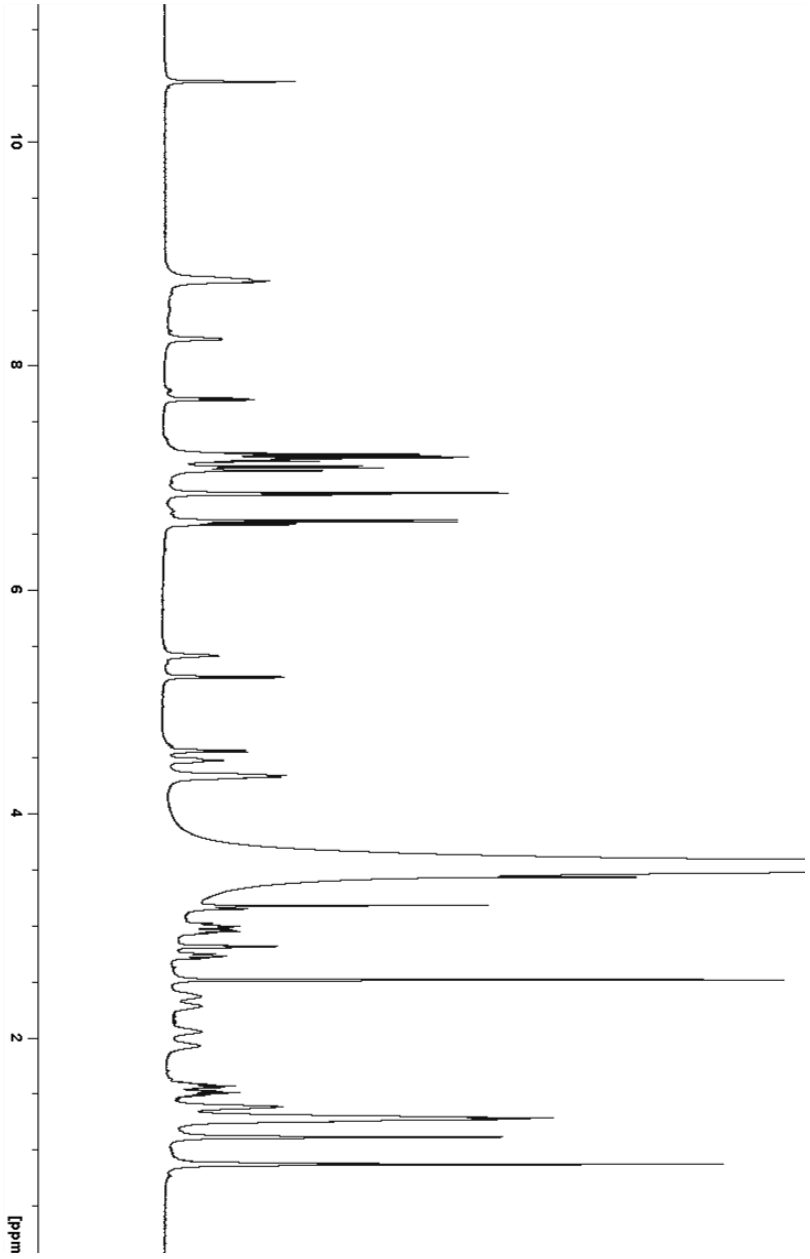


Figure 8.25. ^{1}H NMR (600MHz, DMSO-d_6) spectrum of cystargamide.

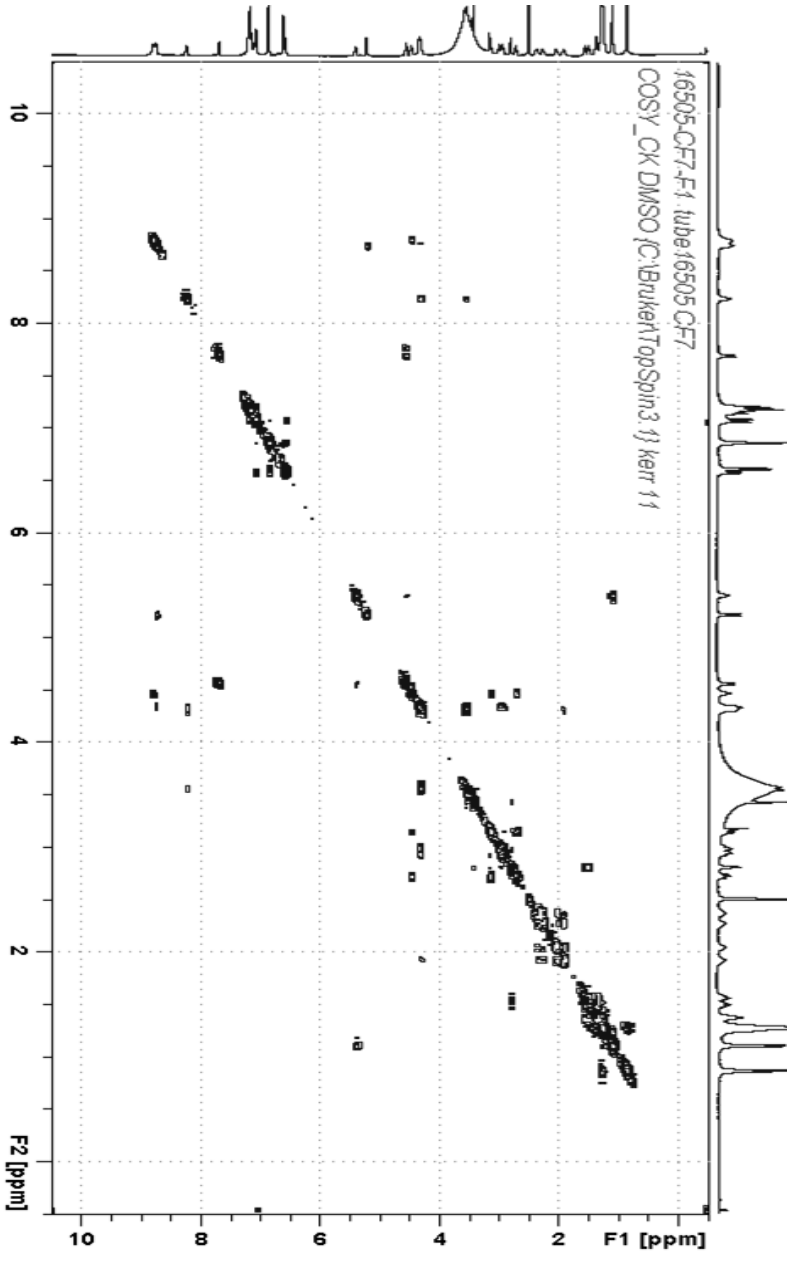


Figure 8.26. COSY NMR spectrum of cystargamide.

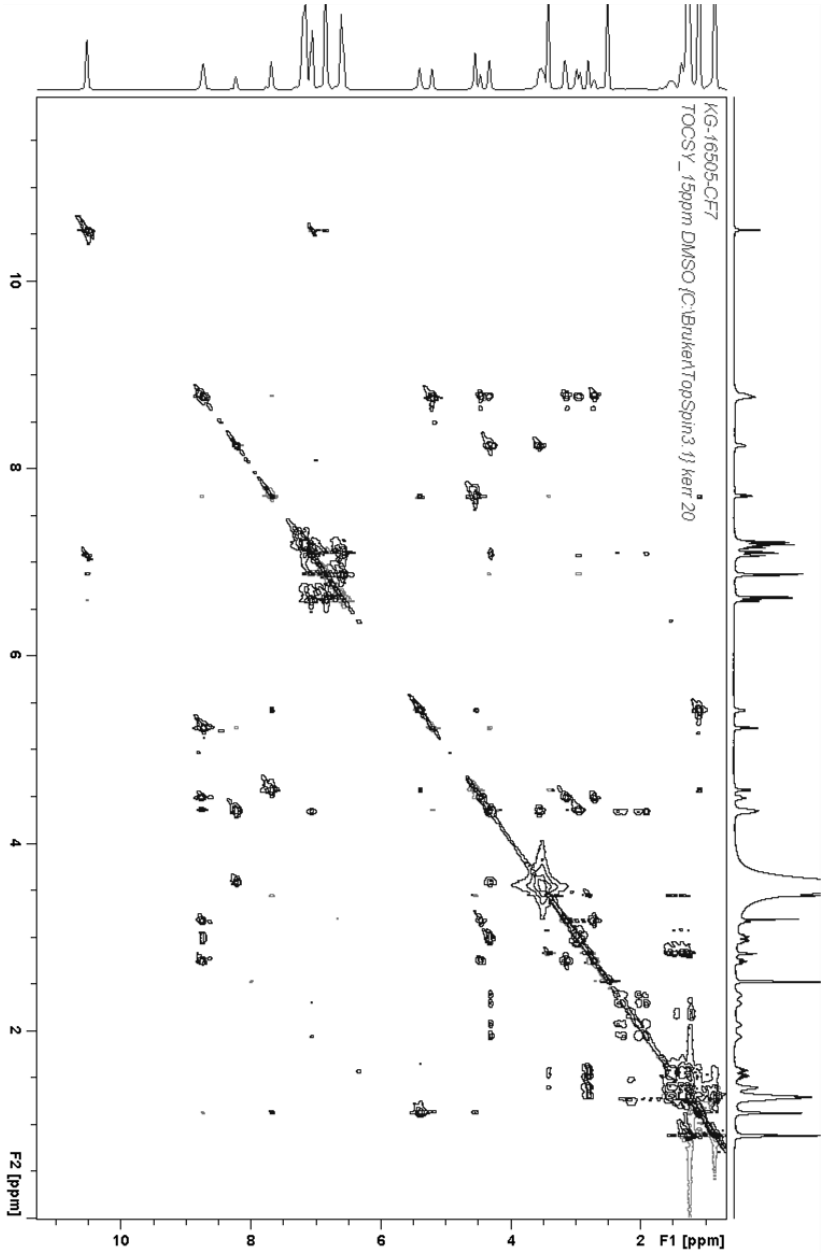


Figure 8.27. TOCSY NMR spectrum of cystargamide.

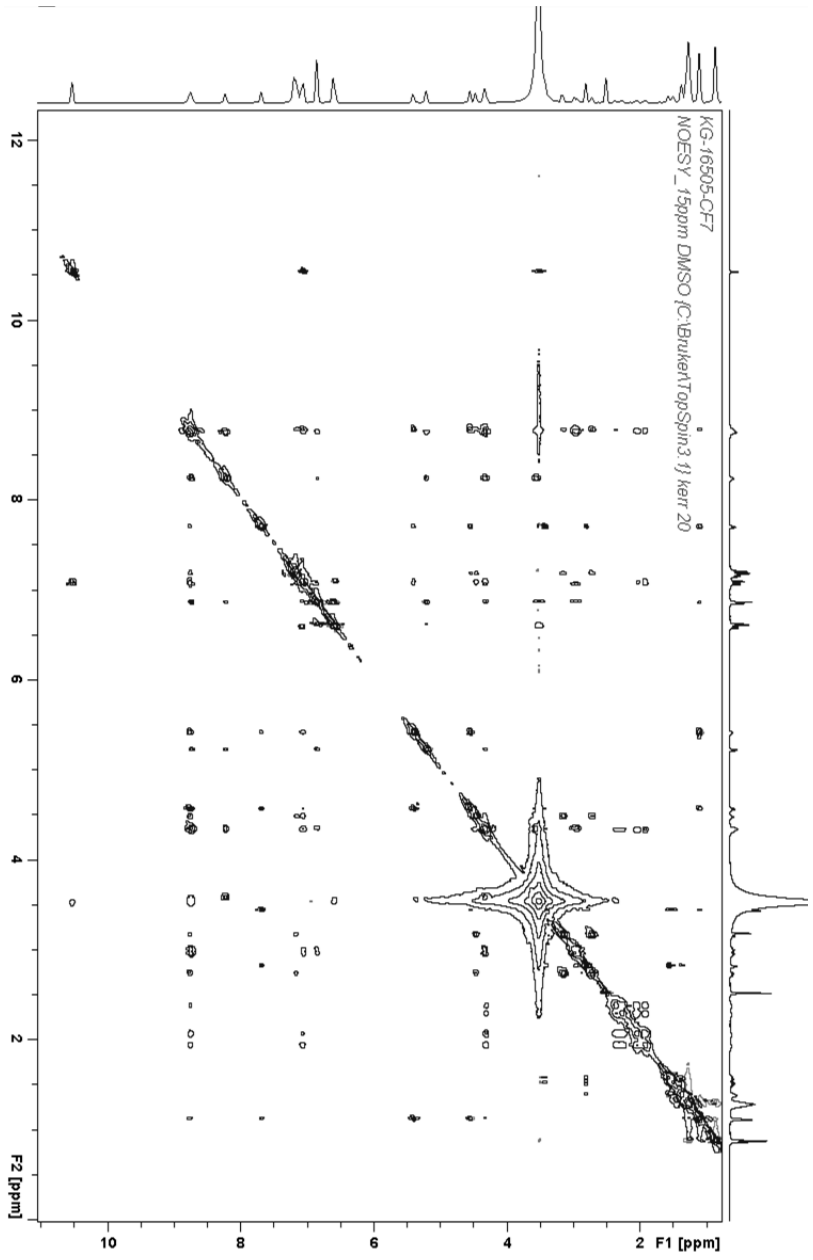


Figure 8.28. NOESY NMR spectrum of cystargamide.

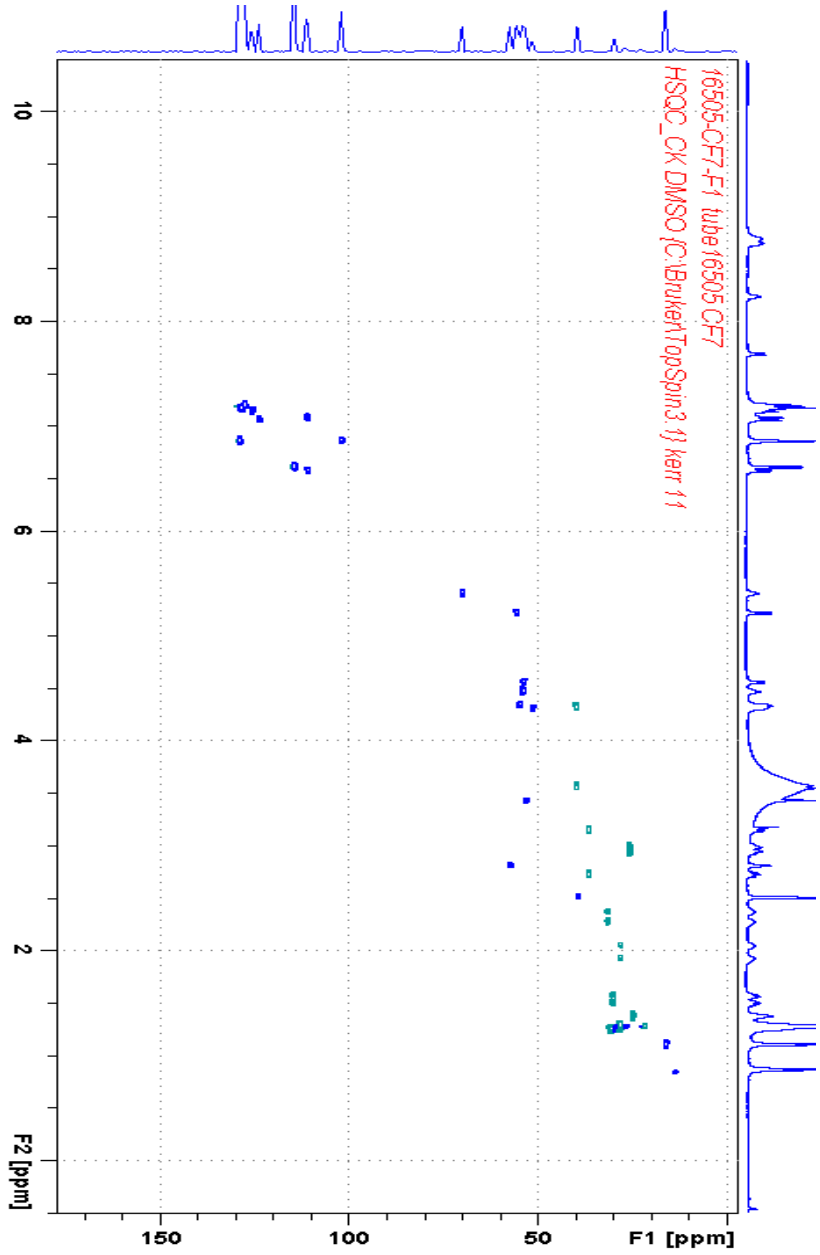


Figure 8.29. HSQC NMR spectrum of cytargamide.

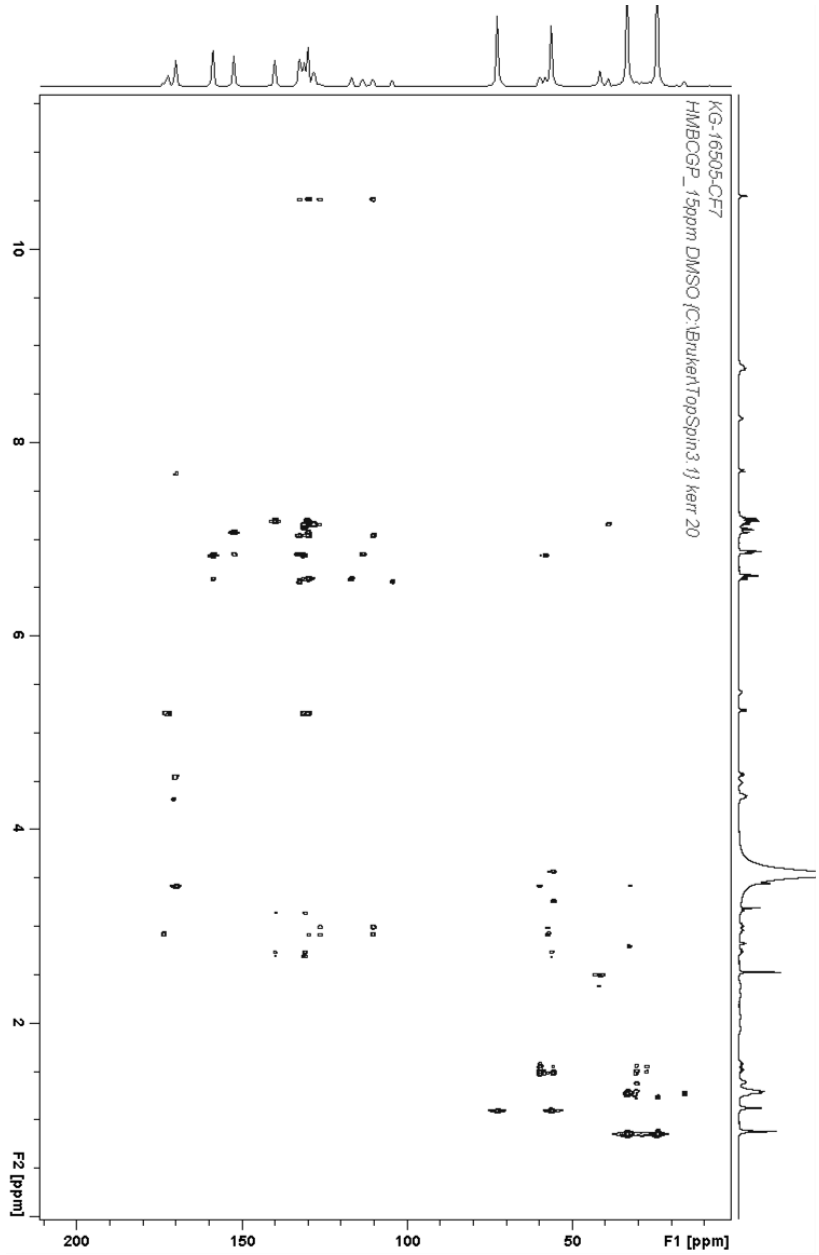


Figure 8.30 HMBC spectrum of cystargamide.

8.4 NMR and MS Spectra Used to Elucidate the Structures of Cystargolides A and B

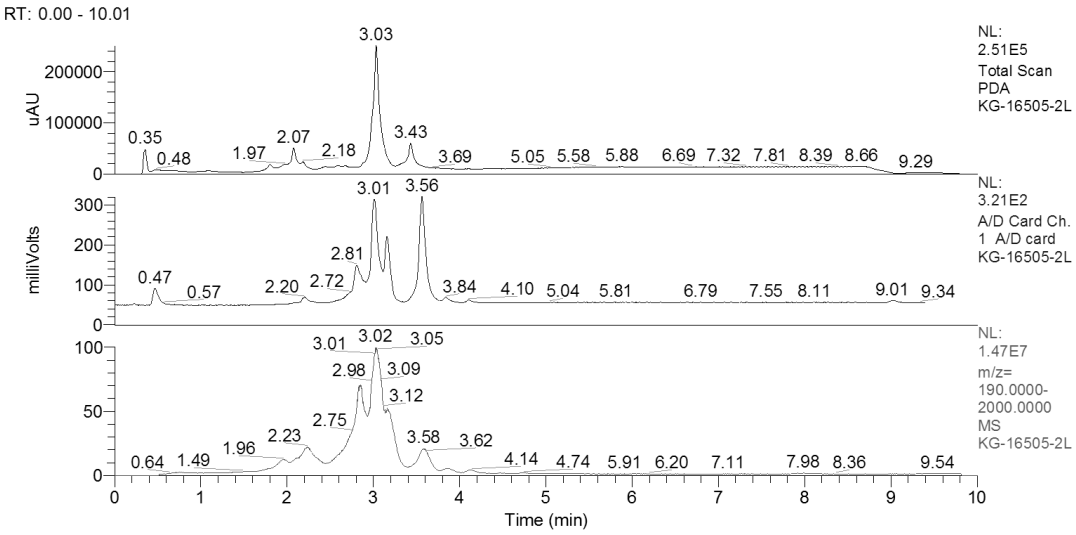


Figure 8.31. Chromatograms of the crude extract from a large scale fermentation of *Kitasatospora cystarginea* grown in 1045 medium.
Three detectors were used to generate 3 chromatograms for the same sample: PDA, ELSD and +ESI-MS.

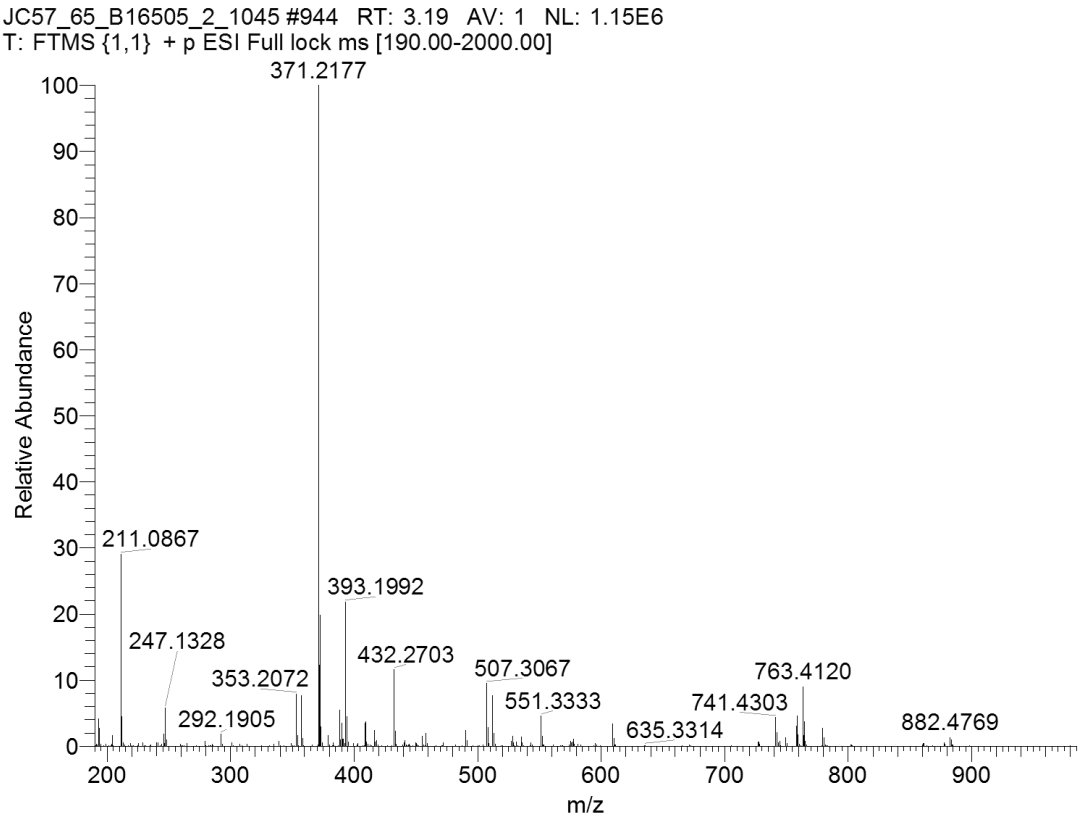


Figure 8.32. +ESI-HRMS of cystargolide A.

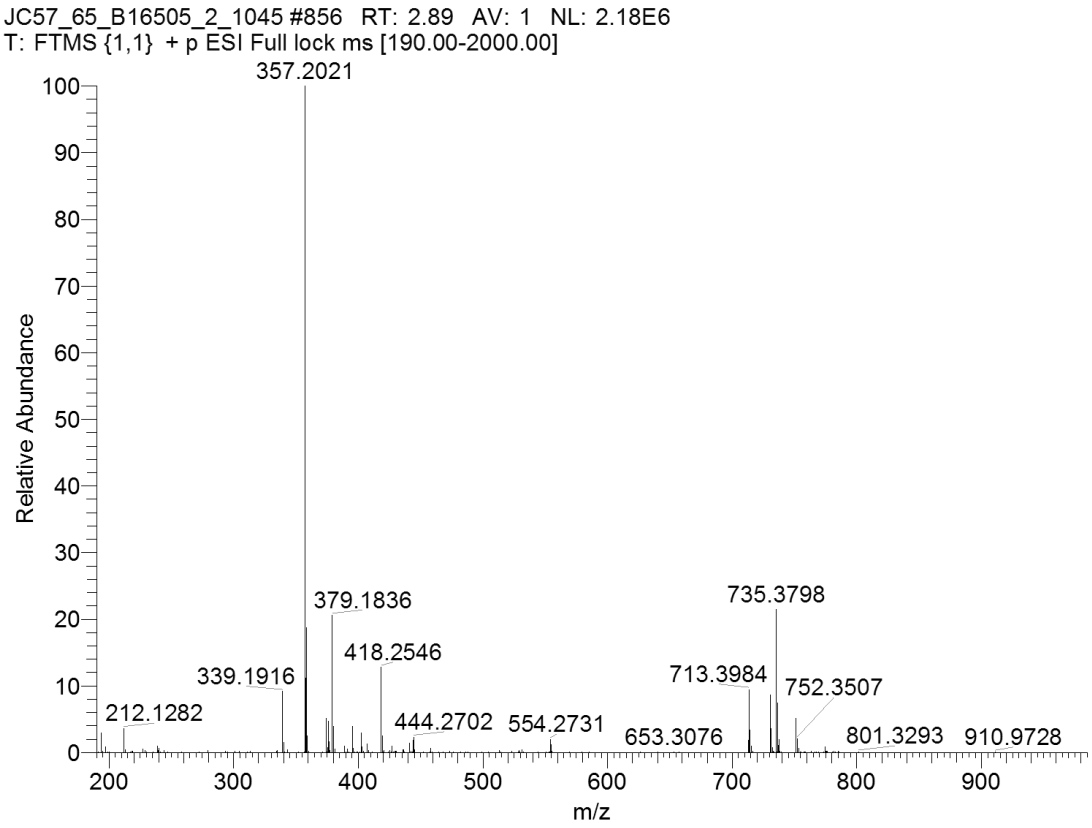


Figure 8.33. +ESI-HRMS of cystargolide B.

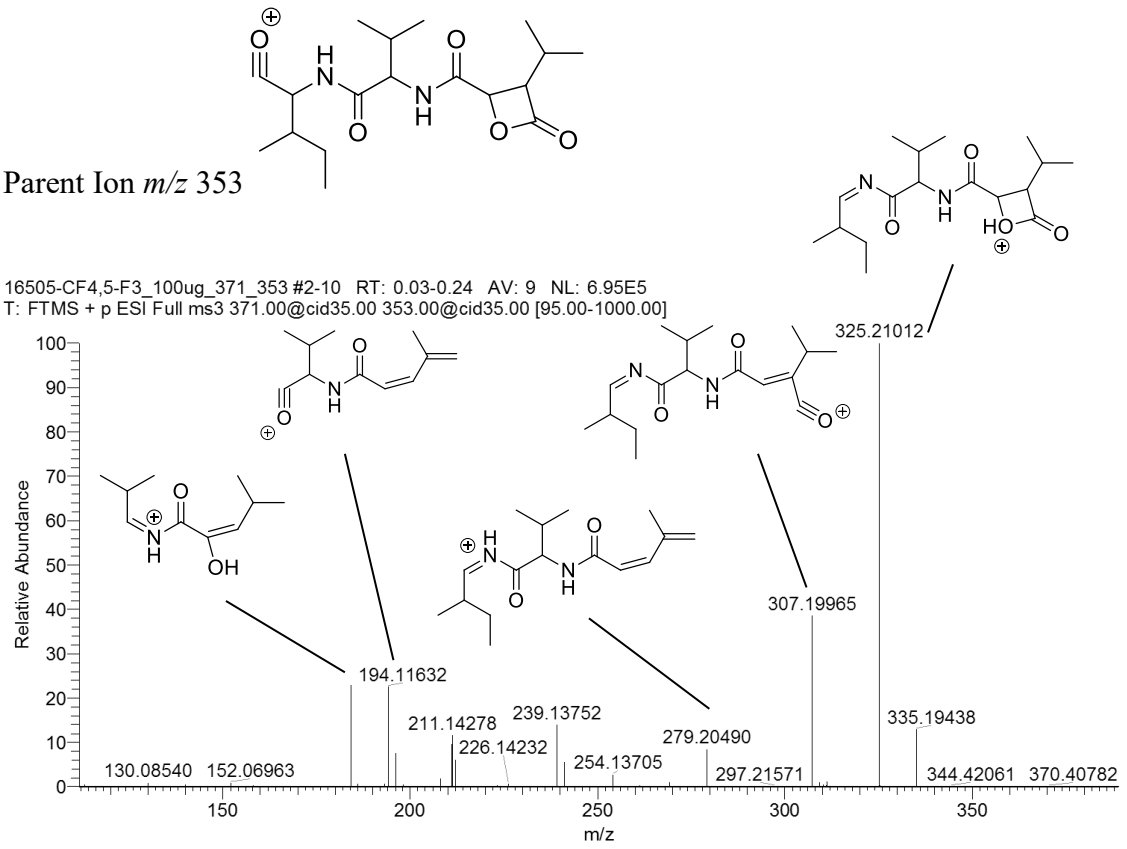
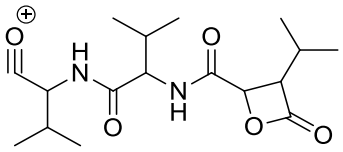


Figure 8.34. MS^3 spectrum of cystargolide A.

Parent Ion *m/z* 339 

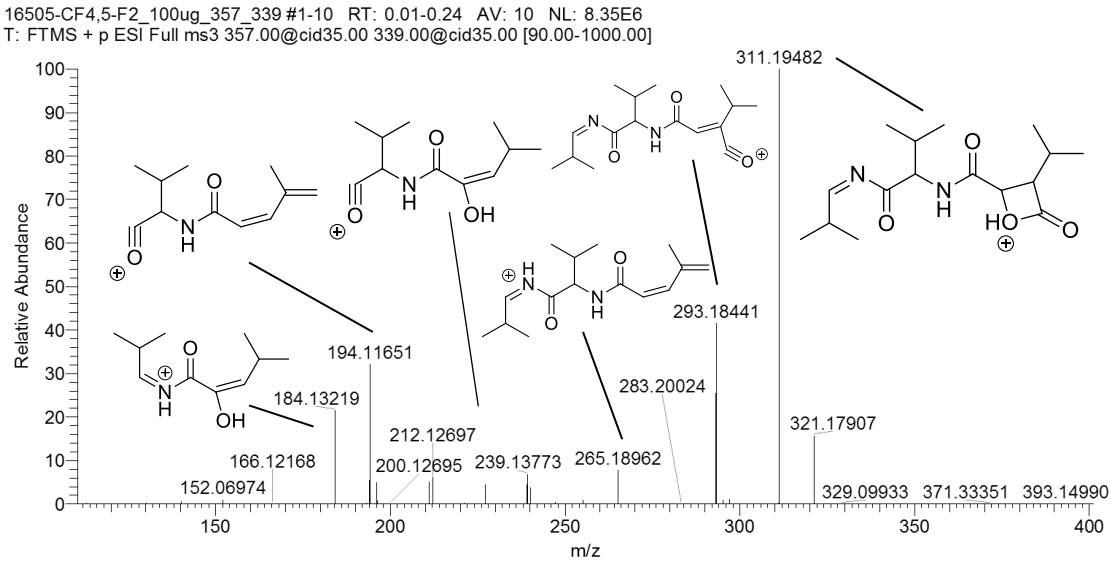


Figure 8.35. MS^3 spectrum of cystargolide B.

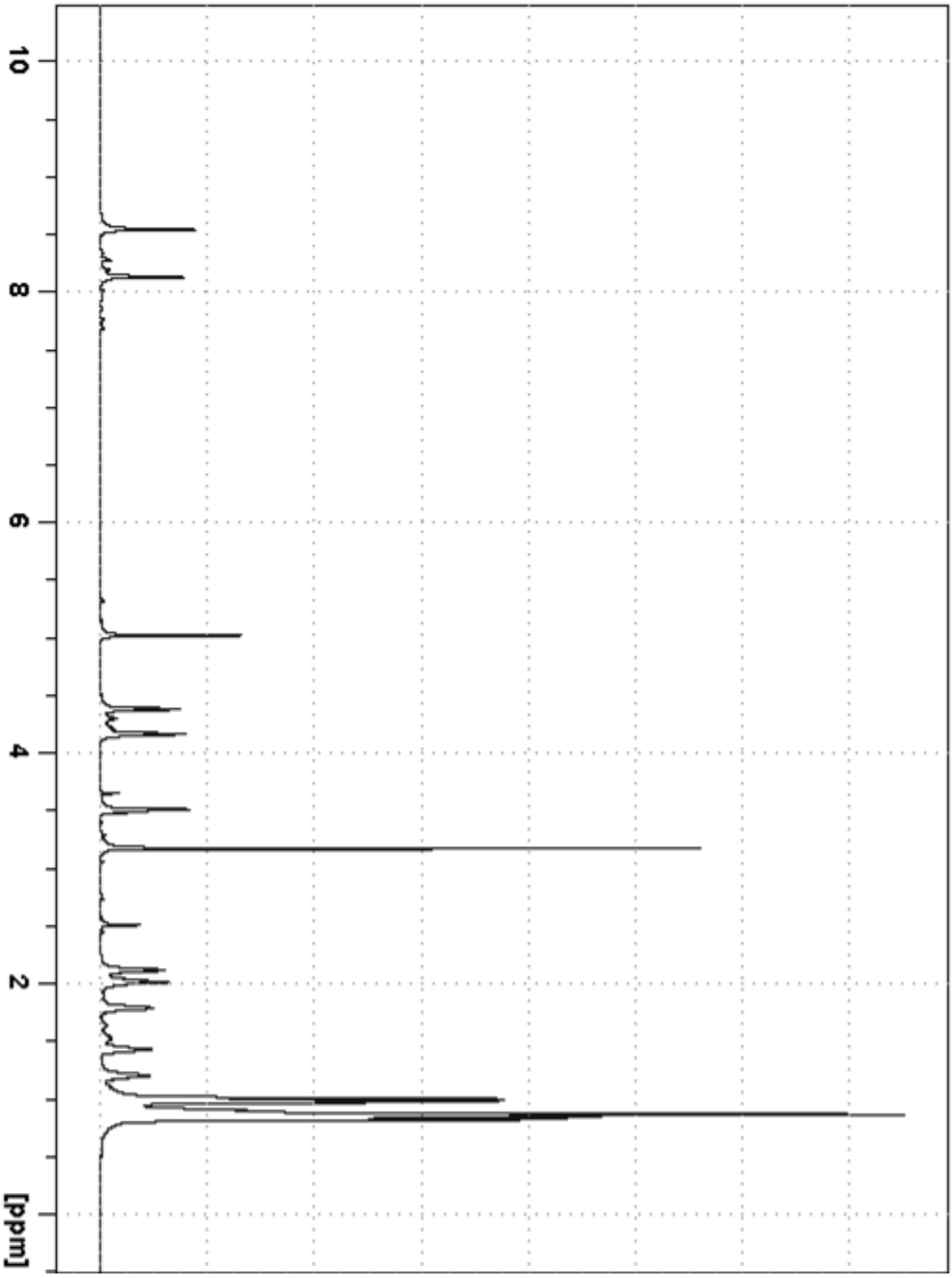


Figure 8.36.^{1}H NMR (600 MHz, DMSO-d_6) spectrum of cystargolide A.

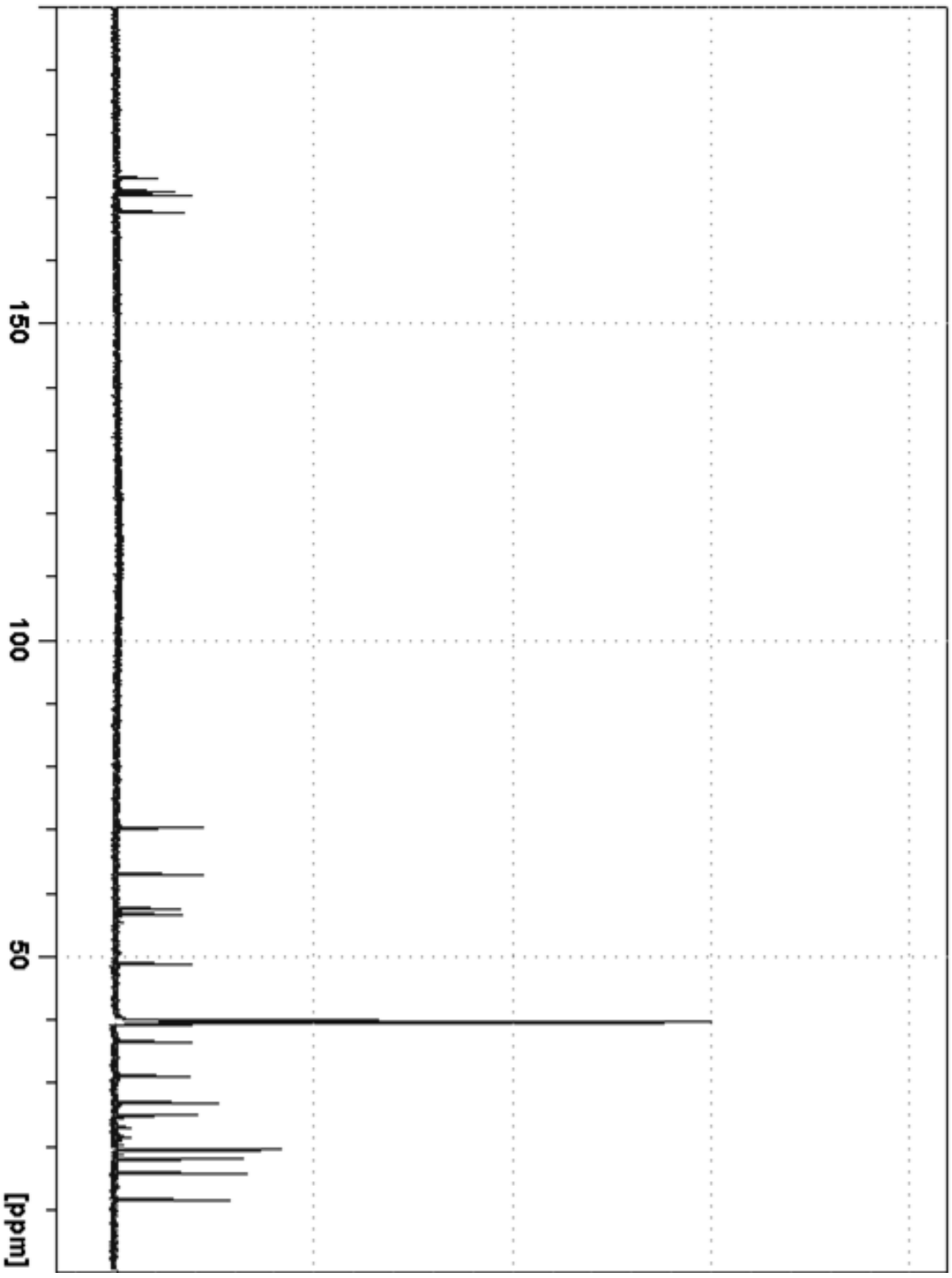


Figure 8.37. ^{13}C NMR (150 MHz, DMSO-d_6) spectrum of cystargolide A.

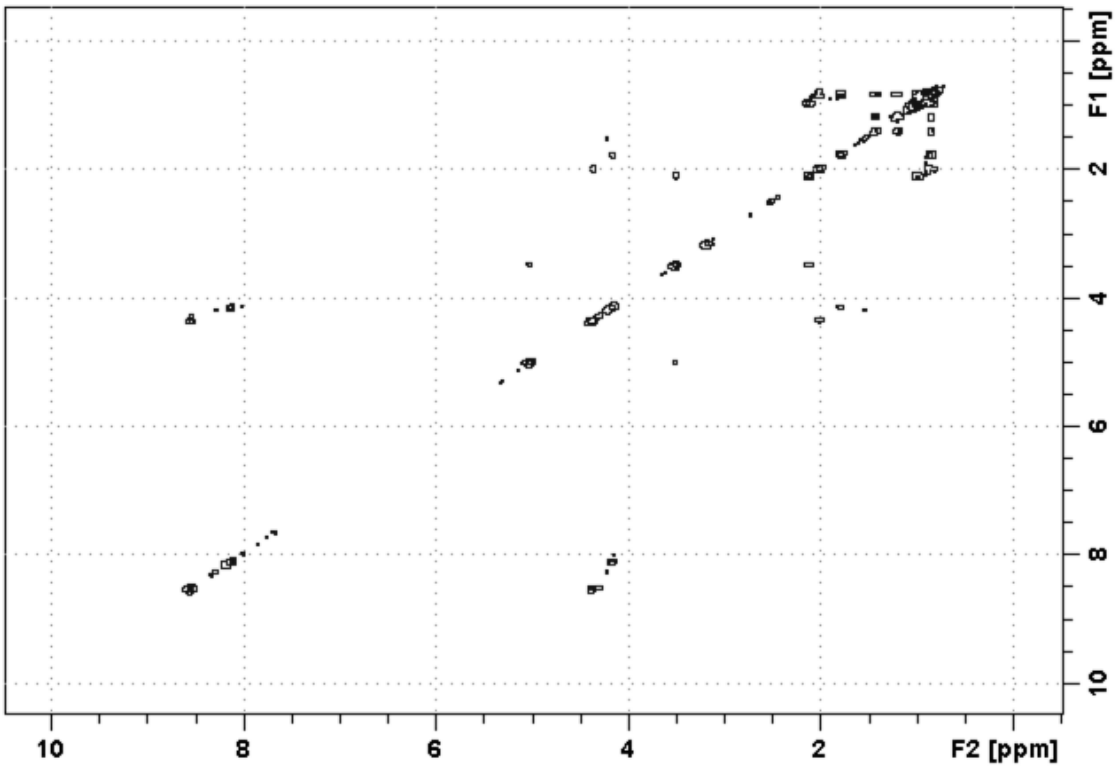


Figure 8.38. COSY NMR (^{1}H, 600 MHz, DMSO-d_6) spectrum of cystargolide A.

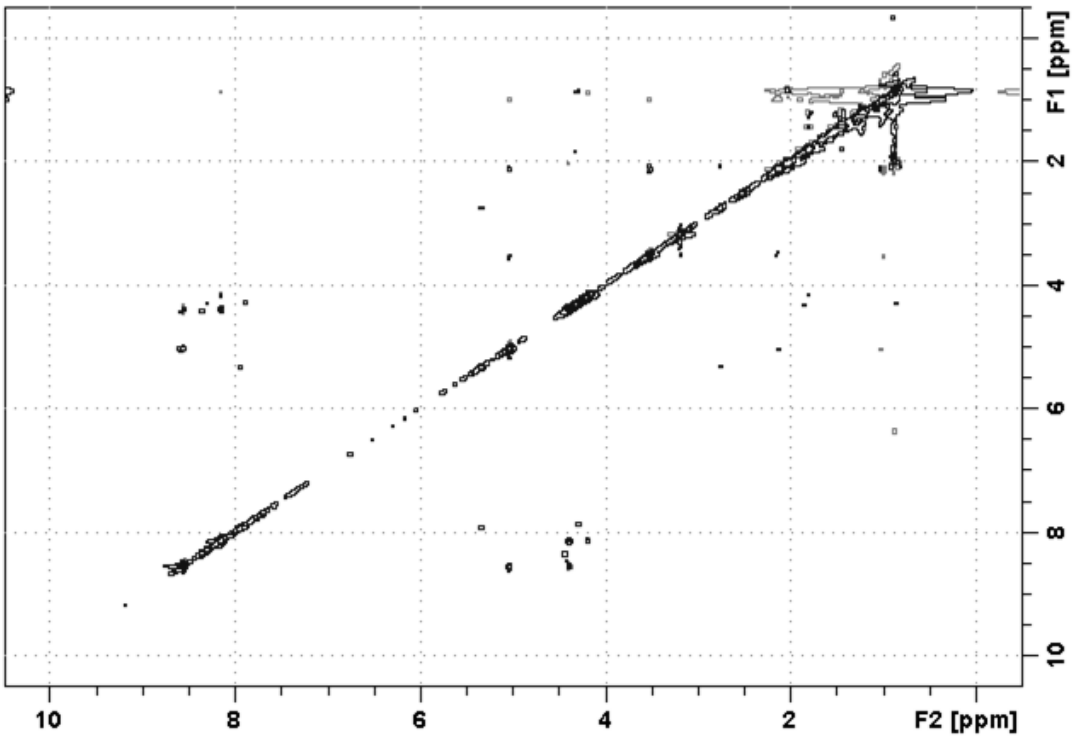


Figure 8.39. NOESY NMR (1H, 600 MHz, DMSO-d_6) spectrum of cystargolide A.

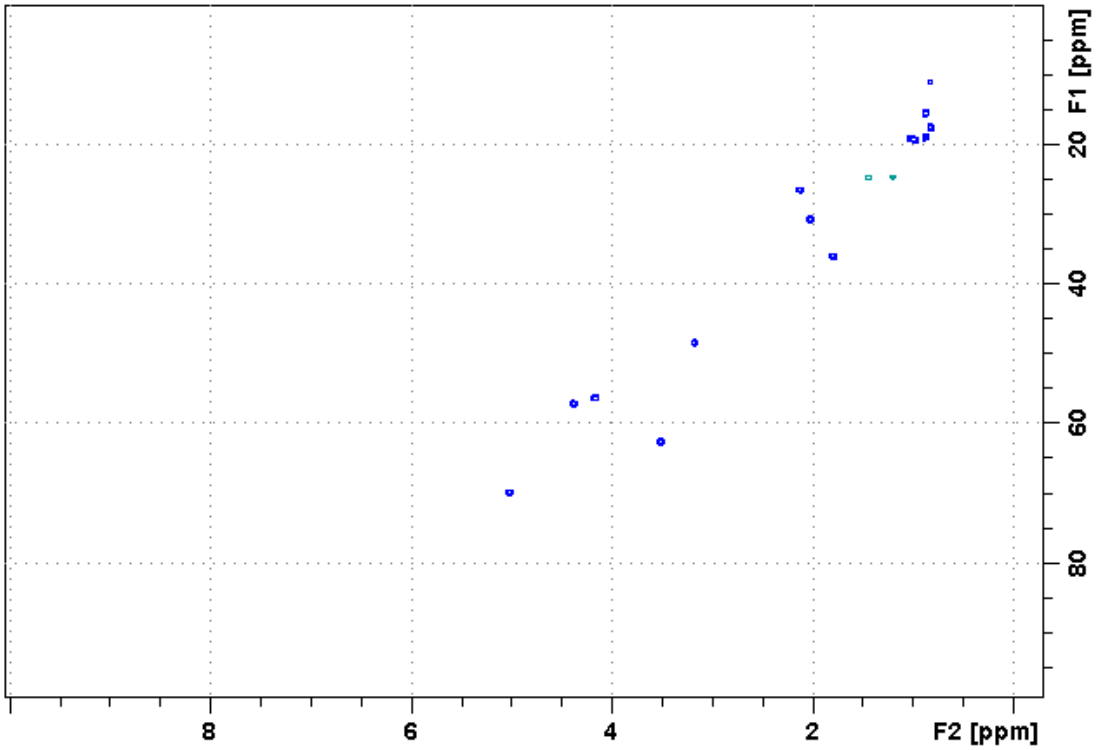


Figure 8.40. HSQC NMR (^{1}H 600 MHz, ^{13}C 150 MHz, DMSO-d_6) spectrum of cystargolide A.

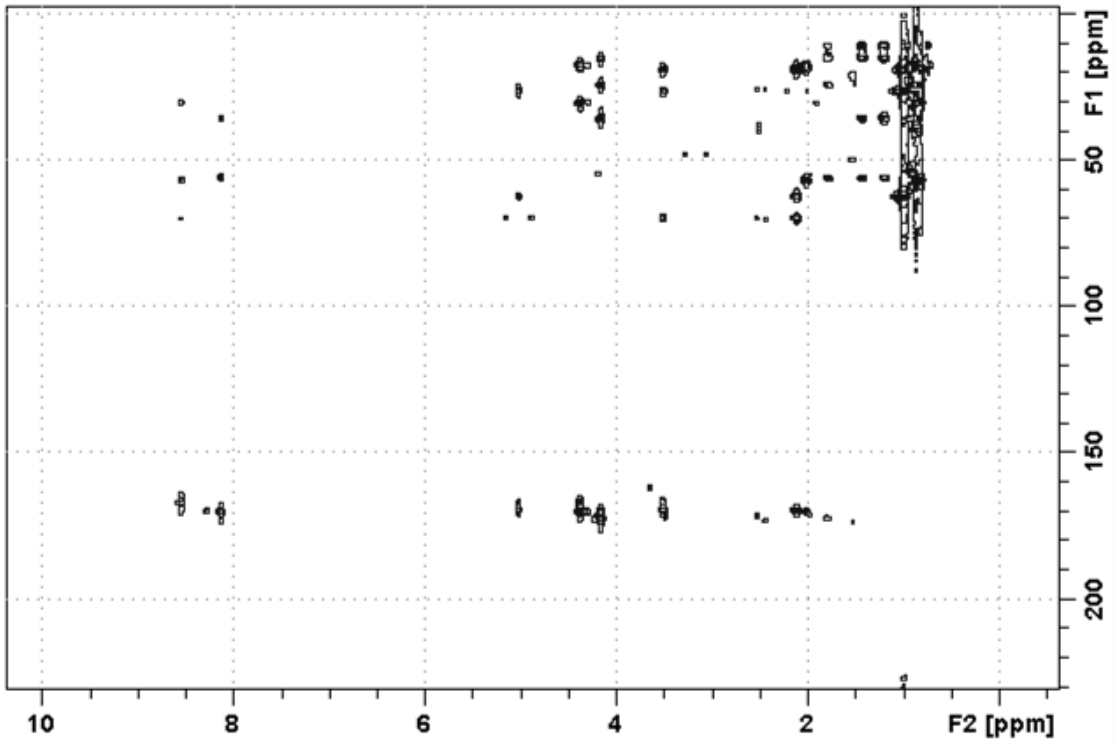


Figure 8.41. HMBC NMR (^{1}H 600 MHz, ^{13}C 150 MHz, DMSO-d_6) spectrum of cystargolide A.

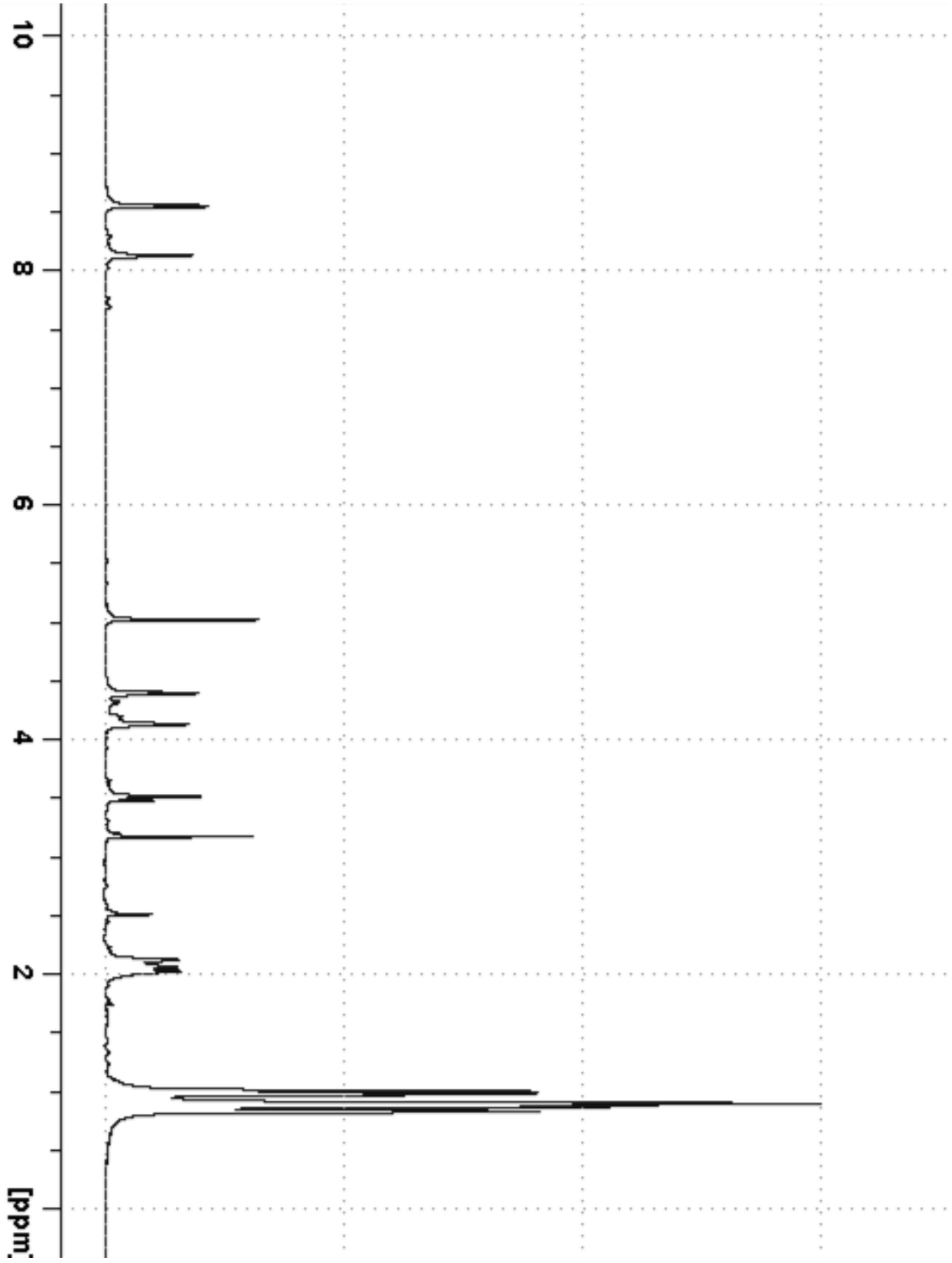


Figure 8.42. ^{1}H NMR (600 MHz, DMSO-d_6) spectrum of cystargolide B

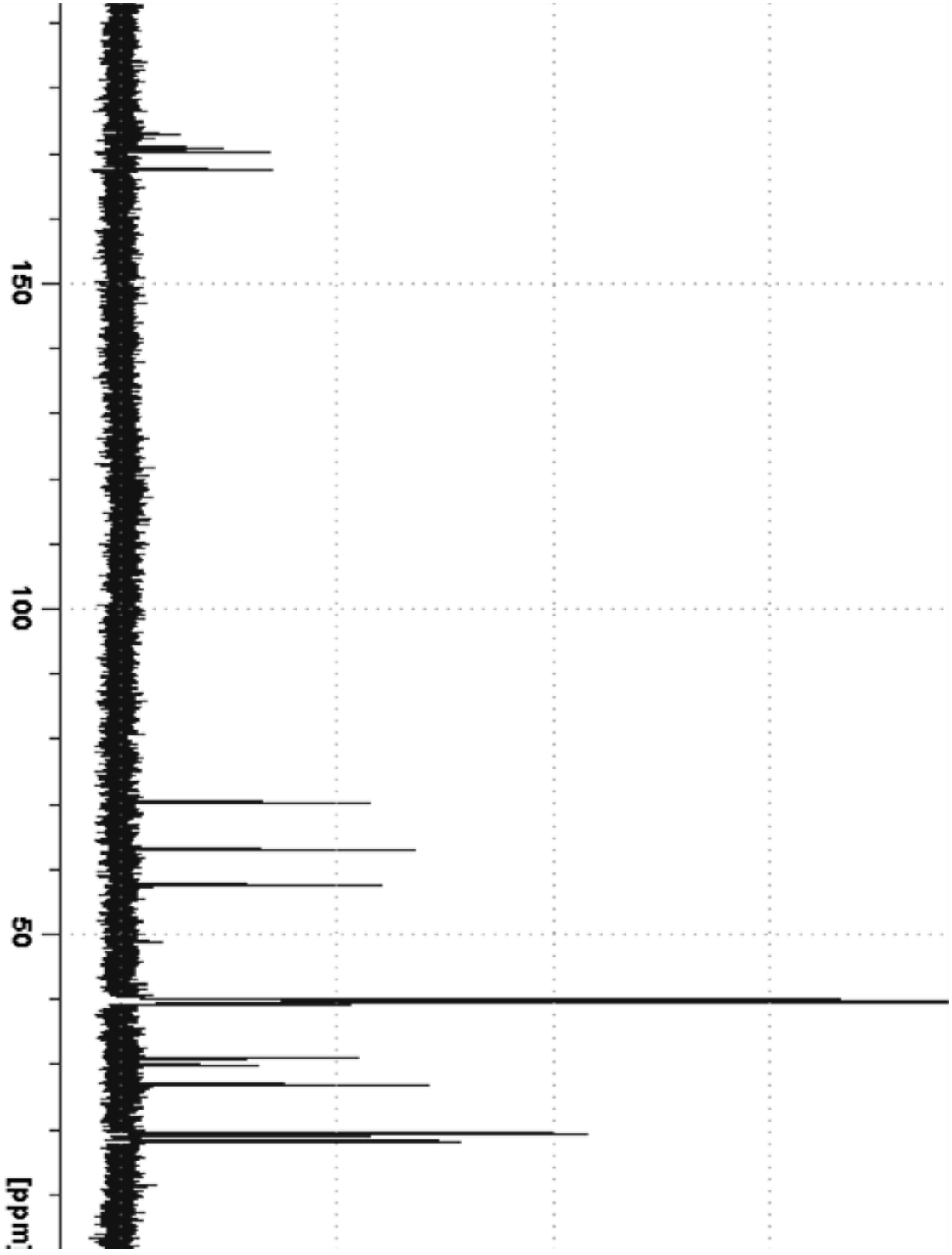


Figure 8.43. ^{13}C NMR (150 MHz, DMSO-d_6) spectrum of cystargolide B.

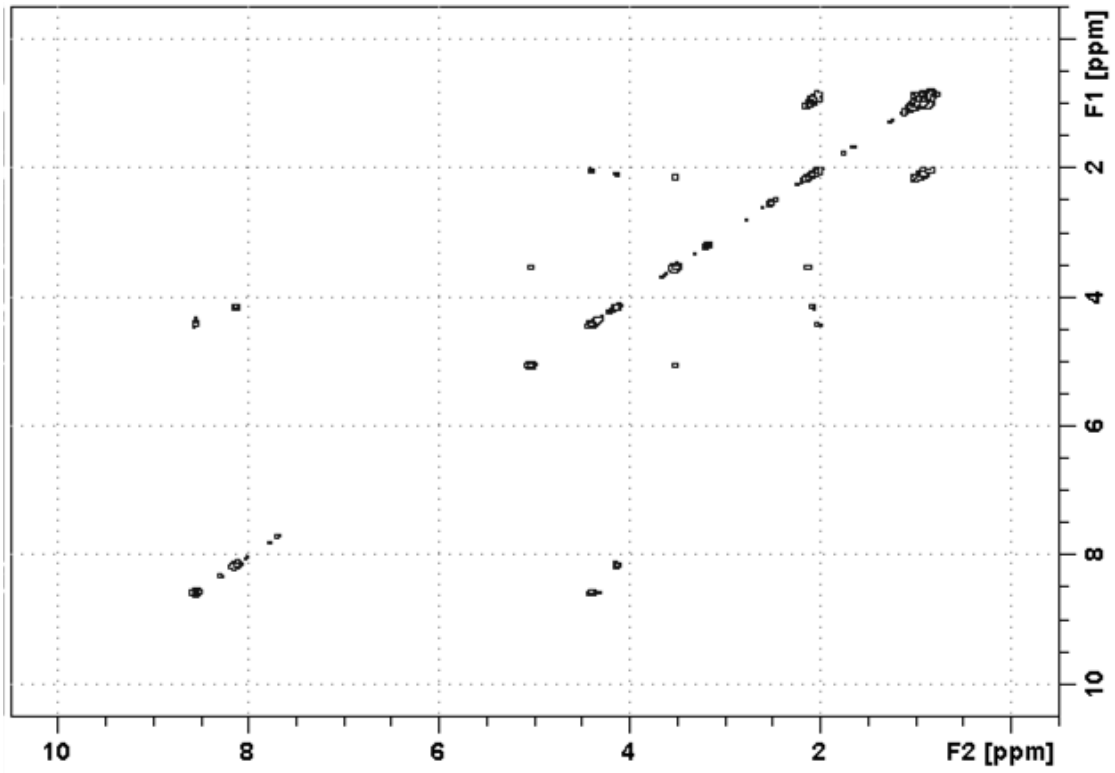


Figure 8.44. COSY NMR (^{1}H, 600 MHz, DMSO-d_6) spectrum of cystargolide B.

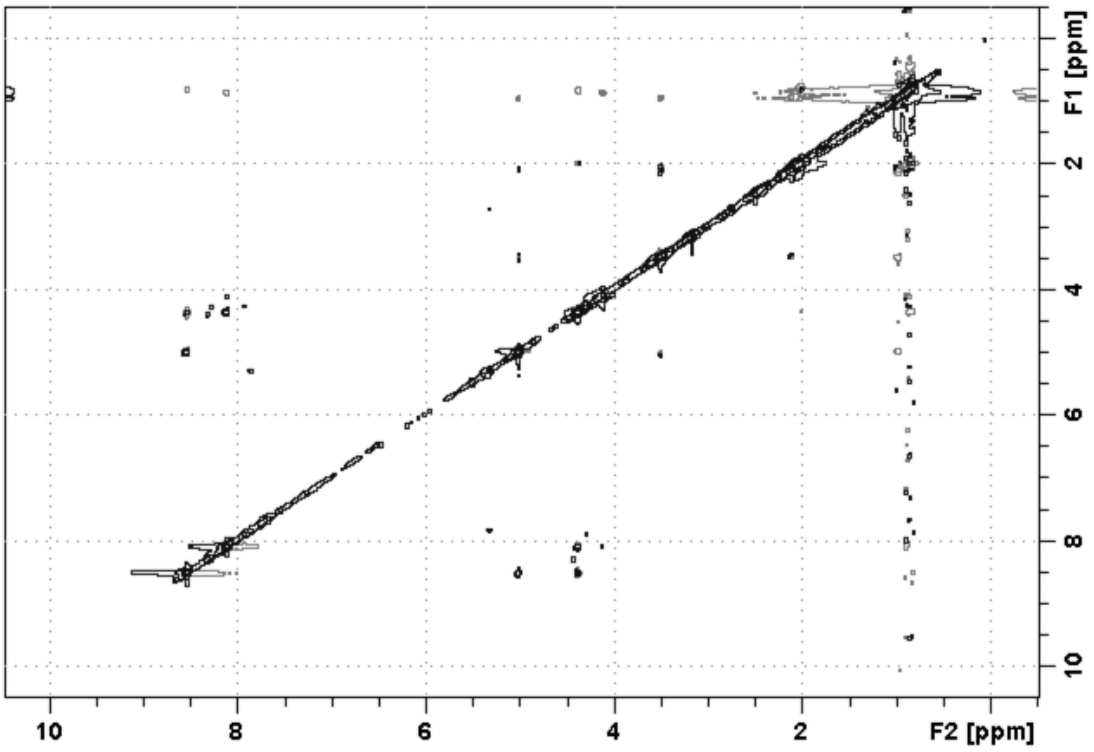


Figure 8.45. NOESY NMR (1H, 600 MHz, DMSO-d_6) spectrum of cystargolide B.

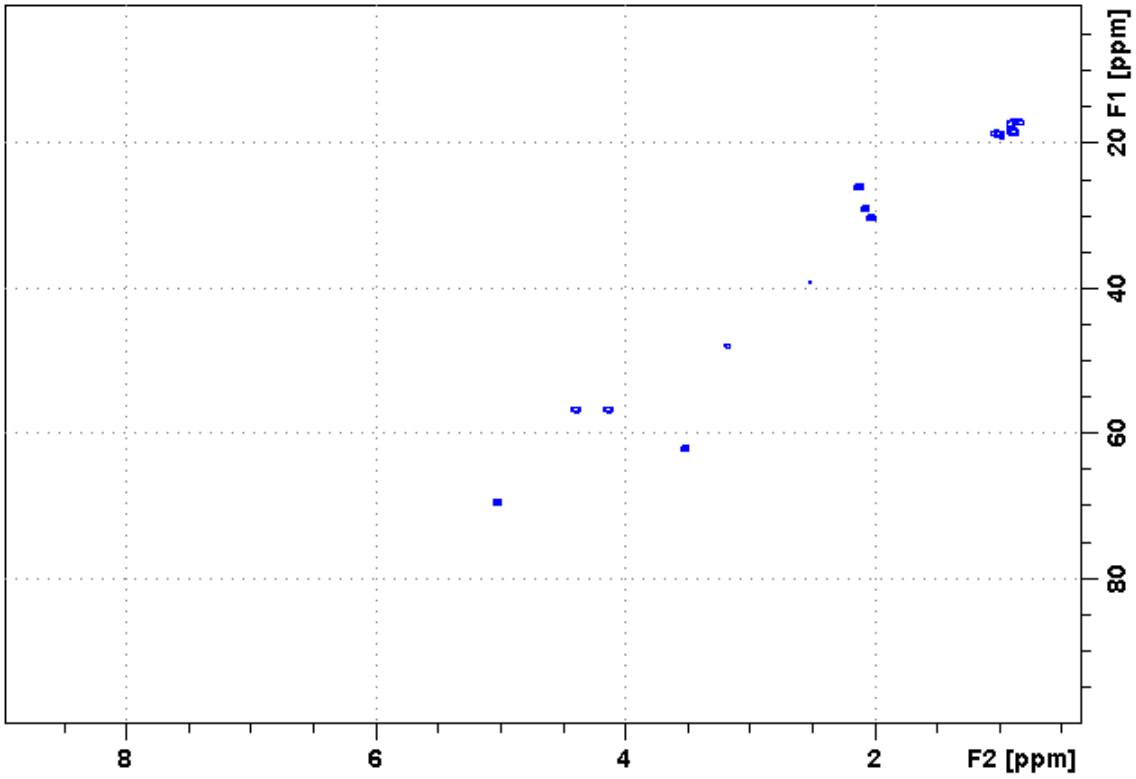


Figure 8.46. HSQC NMR (^{1}H 600 MHz, ^{13}C 150 MHz, DMSO-d_6) spectrum of cystargolide B.

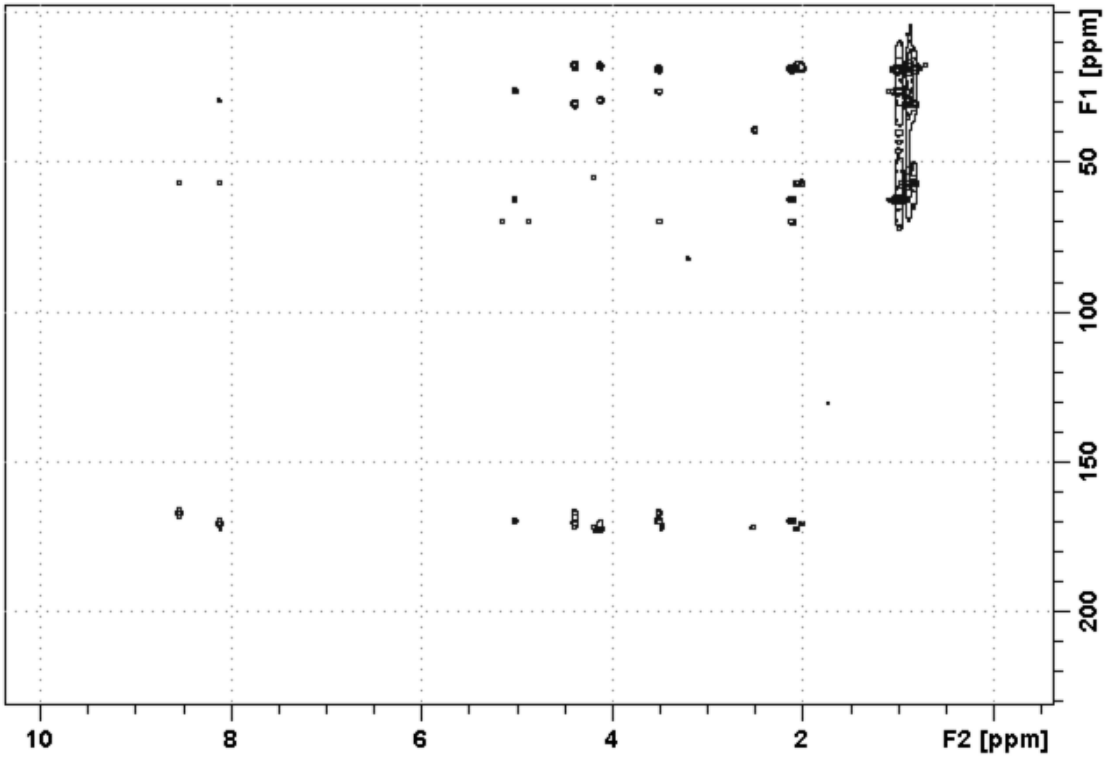


Figure 8.47. HMBC NMR (^{1}H 600 MHz, ^{13}C 150 MHz, DMSO-d_6) spectrum of cystargolide B.

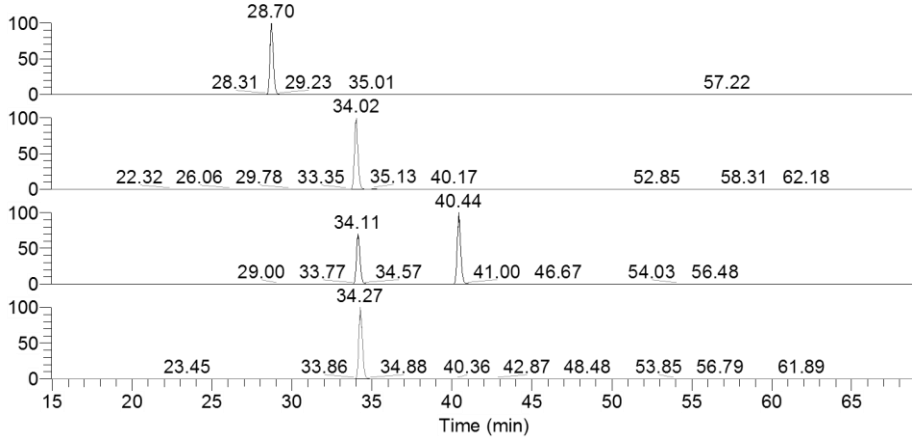


Figure 8.48. Marfey's Analysis. SIM chromatograms of L-FDAA derivatized amino acid standards.

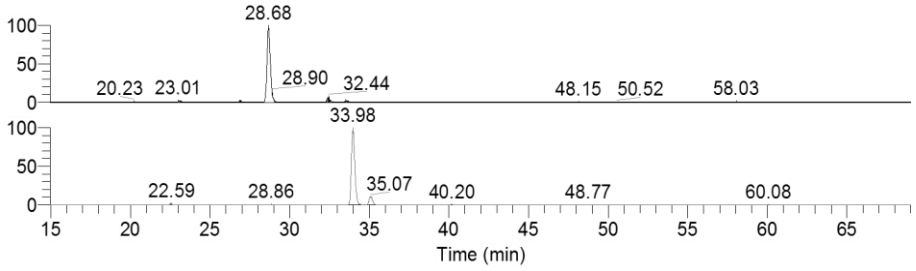


Figure 8.49. Marfey's Analysis. SIM chromatograms of L-FDAA derivatized hydrolysate of cystargolide A.

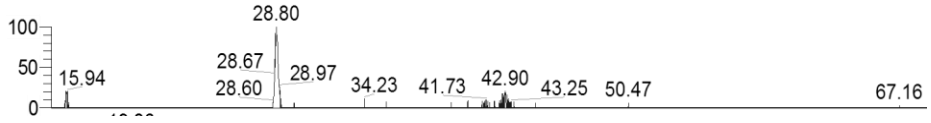


Figure 8.50. Marfey's Analysis. SIM chromatogram of L-FDAA derivatized hydrolysate of cystargolide B.

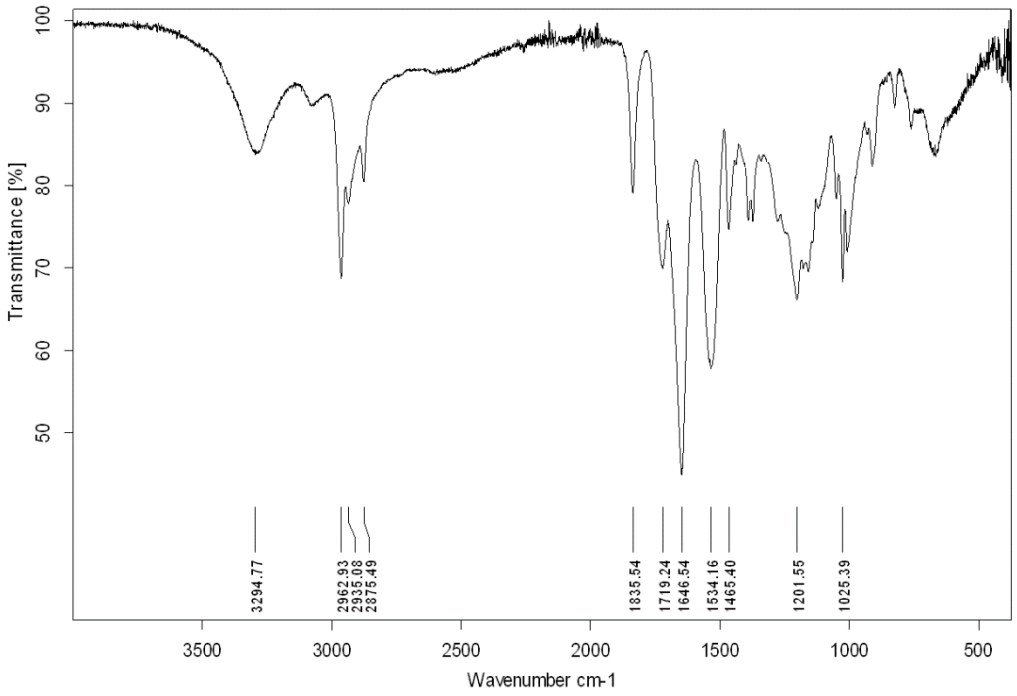


Figure 8.51. IR spectrum of cystargolide A (MeOH, film).

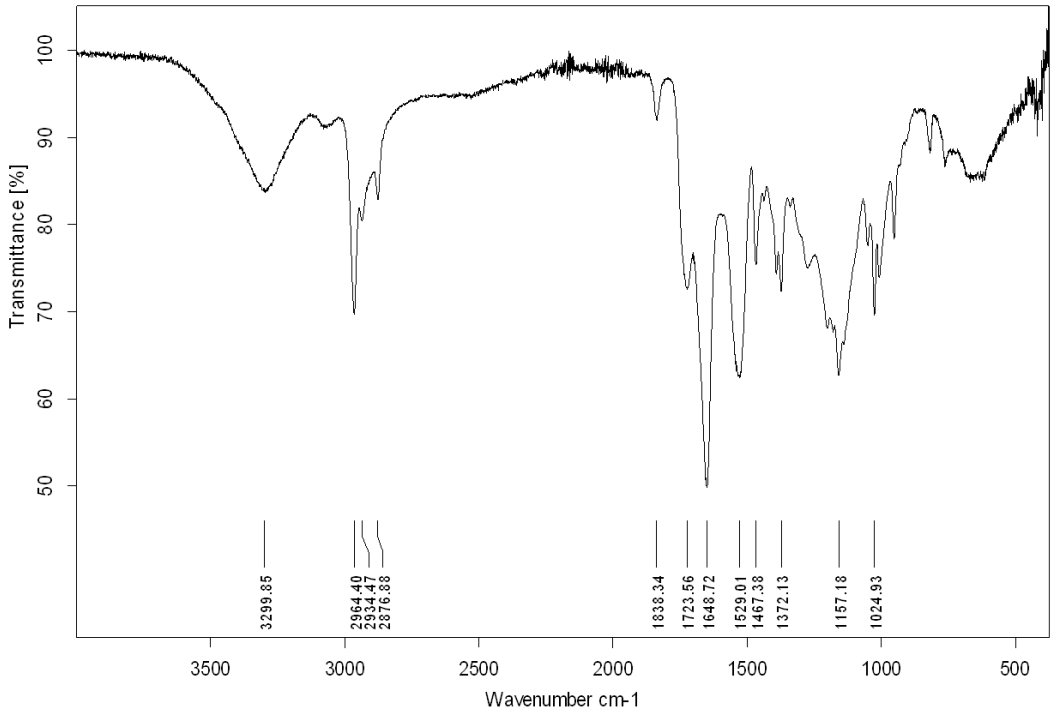


Figure 8.52. IR spectrum of cystargolide B (MeOH, film).